ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

FORECASTING AND NOWCASTING MACROECONOMIC VARIABLES: A METHODOLOGICAL OVERVIEW

Jennifer L. Castle, David F. Hendry and Oleg I. Kitov

Number 674
September 2013

Manor Road Building, Manor Road, Oxford OX1 3UQ

Forecasting and Nowcasting Macroeconomic Variables: A Methodological Overview

Jennifer L. Castle, David F. Hendry and Oleg I. Kitov
Institute for New Economic Thinking at the Oxford Martin School, Oxford University, UK

Abstract

We consider the reasons for nowcasting, how nowcasts can best be achieved, and the use and timing of information. The existence of contemporaneous data such as surveys is a major difference from forecasting, but many of the recent lessons about forecasting remain relevant. Given the extensive disaggregation over variables underlying flash estimates of aggregates, we show that automatic model selection can play a valuable role, especially when location shifts would otherwise induce nowcast failure. Thus, we address nowcasting when location shifts occur, probably with measurement errors. We describe impulse-indicator saturation as a potential solution to such shifts, noting its relation to intercept corrections and to robust methods to avoid systematic nowcast failure. We propose a nowcasting strategy, building models of all disaggregate series by automatic methods, forecasting all variables before the end of each period, testing for shifts as available measures arrive, and adjusting forecasts of cognate missing series if substantive discrepancies are found. An alternative is switching to robust forecasts when breaks are detected. We apply a variant of this strategy to nowcast UK GDP growth, seeking pseudo real-time data availability.

JEL classifications: C52, C51.
KEYWORDS: Nowcasting; Location shifts; Forecasting; Contemporaneous information; *Autometrics*; Impulse-indicator saturation.

Contents

0 Executive summary

Forecasts are defined as made before a period (say a quarter) commences, nowcasts during the relevant period, and flash estimates immediately or shortly after the period ends when disaggregate information remains incomplete. All three activities face similar problems, but the introduction outlines the seven major difficulties that confront successful nowcasting and the calculation of accurate flash estimates. The first four difficulties relate to the disaggregates required to construct a flash estimate of a relevant aggregate (such as GDP), and comprise: a 'missing data problem' as not all disaggregated contemporaneous data are known when needed; a 'latency problem' as data on some variables arrive long after they are needed; but rarely the same variables on each occasion, leading to 'a changing database problem'; and even when 'known', disaggregates may be revised later, the 'measurement error problem'. These difficulties are exacerbated by the occurrence of unanticipated location shifts, namely changes in the long-run means of the variables being nowcast, which we call the 'breaks problem'. In addition, 'continuous high-frequency updates' of aggregate measures require handling large data sets, potentially creating a 'problem of more variables, N, than observations, T', and facing a 'non-synchronous release problem', all of which necessitate a flexible software system.

To tackle these problems, we propose a strategy that involves building models of all N disaggregate series by automatic model selection methods, forecasting all N variables before the end of each period, testing for shifts in the available time series as their data arrive, and either adjusting flash estimates of missing variables if shifts are detected in cognate measured series, or switching to robust forecasts when breaks are detected. To explain the basis of our proposal, we first review some of the fundamental issues that affect forecasting. This entails clarifying why conditional expectations derived from estimated models given available information need not produce good forecasts when unanticipated location shifts occur. Moreover, while more information will generally lower unexplained variances, less information by itself does not lead to major problems, unless it is information about otherwise unanticipated shifts.

As such lessons are relevant to nowcasting, we review the ten key difficulties in forecasting using a scalar autoregressive-distributed lag equation where the unmodeled variable is known in the 'future' period. The resulting complex set of forecast-error outcomes can be explained by the presence or absence of location shifts: when such shifts occur, forecast failure results irrespective of the goodness or otherwise of the in-sample model; when location shifts do not occur, there is no obvious failure almost irrespective of how poor the model is and how many parameters of the data-generating process (DGP) shift, and by how much. We next show that robust forecasting devices may forecast better than any structural model in such shifting processes, as measured by root mean-square forecast errors, illustrated by three empirical applications. Similar problems beset nowcasting, and flash estimates so have related solutions.

The huge number of variables potentially involved in our approach highlights the need for automatic model selection methods capable of efficiently handling more variables than observations while tackling multiple location shifts. Consequently, we briefly describe how *Autometrics* does so, using impulse-indicator saturation (IIS) to handle in-sample location shifts. After modeling one series, the computer code can be generated in Ox, and 'looped' to apply to automatically modeling many other series, providing one flexible software solution.

Readers familiar with our research on forecasting in the face of location shifts or automatic modeling can go directly to the latter part of our chapter on using such findings to improve the production of 'good' nowcasts by nowcasting aggregates from disaggregates. The key difference between forecasting and nowcasting is the availability of contemporaneous data, so we discuss the possible roles of surveys and other covariate information, Google Trends, and prediction market data. We illustrate the approach for *ex post* nowcasts of UK GDP growth, but seeking pseudo real-time data availability.

1 Introduction

Nowcasting is 'forecasting' the current or recent aggregate state of an economy.[1] It can be undertaken either at the initial stage of improving 'flash' estimates within a National Statistical Agency, which would otherwise just be based on inferences from the intermittent arrival of highly disaggregate data measures, or by an outside organization using higher-frequency available information such as surveys to improve available aggregates, or both. Here we focus mainly on the former, and explain the use of recent developments in both modeling and forecasting to improve the accuracy and timeliness of initial data releases. Approaches based on ascertaining the state of an economy from a wide range of information sources, so 'high-frequency estimates' of relevant aggregates are always available, are discussed by Angelini, Camba-Méndez, Giannone, Rünstler and Reichlin (2011) and Giannone, Reichlin and Small (2008), with a survey in Bánbura, Giannone and Reichlin (2011). Some of the issues and methods discussed here apply to both approaches. We refer to any procedure that uses additional information when producing contemporaneous aggregate data, beyond just cumulating observed disaggregates, as nowcasting.

There are obvious practical reasons why up-to-date information is needed on economic aggregates, especially those which play a key role in economic policy decisions. There are four difficulties confronting the production of timely and accurate low-frequency aggregates, necessitating nowcasting, as we now explain. A fifth problem (denoted the 'breaks problem' below) can adversely affect the accuracy of all forms of nowcasts. There are two other difficulties facing the production of 'continuous high-frequency estimates' of aggregates. We address these seven issues in turn.

First, not all disaggregated contemporaneous data are available at the time their information is required to construct a relevant aggregate–the **'missing data problem'**. Aggregates cannot be constructed just by adding up the relevant observed disaggregates at the end of the relevant period. For preliminary estimates of UK GDP, for example, the information is only available up to 24 days after the end of the quarter for which a flash estimate is required (see Reed, 2000, 2002). This is the central basis for 'nowcasting' at any Statistical Agency, essentially forecasting the missing disaggregates.[2]

Secondly, a **'latency problem'** is ever present, in that some variables have to be estimated or 'infilled' even long after the end of the period just noted. A 'nowcast' of those disaggregates could lead to a more timely and accurate aggregate measure.

Thirdly, different components of the data entering the aggregates for which nowcasts are required are unavailable in different periods (see e.g., Clements and Hendry, 2003), and so are missing on a non-systematic basis. Consequently, a consistent subset of information is rarely available, inducing a **'changing database problem'**, which is bound to affect the design of any system for nowcasting.

Fourthly, many disaggregate economic time series are themselves only preliminary estimates, which are subject to potentially substantial revisions in due course as more information accrues, so are not necessarily a reliable and accurate guide to current conditions. Faust, Rogers and Wright (2007) report that revisions to GDP announcements are quite large in all G7 countries, and are fairly predictable in the UK, mainly because of reversion to the mean, which they interpret as due to removing measurement 'noise'. Thus, **'measurement error problems'** are also pervasive, and consequently, even the 'observed data' cannot be taken at face value, and may be no more accurate than nowcast values.

[1]We are grateful to Michael P. Clements, Neil R. Ericsson and Nicholas W.P. Fawcett for helpful contributions to the underlying research, which was supported in part by grants from the Open Society Foundations and the Oxford Martin School.

[2]A separate 'missing data problem' is that the existence of some products, companies, etc., may be unknown to the National Statistical Agency at the time, and only incorporated into the data base much later: we do not address that issue here. This is also an extreme version of the 'measurement error problem' discussed in §6.8.

Fifthly, a **'breaks problem'** from unanticipated location shifts (changes in the previous means of the variables under analysis) always threatens to disrupt the accuracy of nowcasts, whether based on 'in-filling' for missing disaggregate data, or the production of 'continuous high-frequency measures' of aggregates. Careful attention is required to avoid shifts contaminating flash estimates. In §6, we propose that 1-period ahead forecasts of the disaggregates should always be produced, enabling immediate evaluation against incoming measured outcomes, to act as an 'early warning signal' when measured series and forecasts depart radically. Comparisons would facilitate rapid action to be taken following any deterioration in the performance of a nowcasting model or method, or warn of measurement problems in the 'actual' series. A cross check could be based on robust forecasts as dicussed in §4.3, and this applies to both nowcasting approaches.

Next, to produce 'continuous high-frequency updates' of aggregate measures requires handling a potentially vast amount of data, which we denote the **'problem of more variables, N, than observations, T'**. In fact, this difficulty can also apply to methods aimed at improving aggregate flash estimates directly. Possible solutions include data reduction (using automatic model selection), data combination (as in principal components or factor methods), and model combination (as in model averaging).

Last, some of the high-frequency data to be used to estimate the low-frquency aggregate are not released in a systematic or synchronous way, and may have different publication lags with each release. Like the 'changing database problem', this **'non-synchronous release problem'** necessitates a flexible software system.

The interaction of these seven problems poses major uncertainties for nowcasters. At first sight, the difficulties are similar to those facing forecasters, and indeed we will draw on the literature analyzing forecasting during breaks. However, a key difference is the presence of contemporaneous observations–which logically could never occur in a forecasting context–and there are many possible sources of such evidence, including rapidly and frequently observed data on variables such as retail sales and financial markets; correlates like road traffic and air passenger numbers or energy consumption; surveys; and more recent innovations, including Google Trends (see Choi and Varian, 2012), mobile phone data, and prediction markets (see Wolfers and Zitzewitz, 2004).

Castle, Fawcett and Hendry (2009) and Castle and Hendry (2010) propose a framework for nowcasting that addresses the issues of shortening the 'latency', in-filling for 'missing data' despite ongoing shifts, reducing 'measurement errors', while being sufficiently flexible to handle 'changing databases' and 'non-synchronous releases' yet tackling the 'breaks problem'. Their approach exploits many sources of contemporaneous data in an automatic model selection algorithm designed to handle $N > T$, as well as tackle unanticipated location shifts both directly and by using robust forecasting devices. We explain the various ingredients of their approach, and illustrate by an application, building on Ferrara, Guegan and Rakotomarolahy (2010). Despite their differences, there are important lessons from the current theory and practice of forecasting relevant to nowcasting, so we first reconsider forecasting, once the orphan of econometrics, but now a topic of major research, with extensive treatments in the four recent *Handbooks* by Clements and Hendry (2002), Elliott, Granger and Timmermann (2006), Rapach and Wohar (2008), and Clements and Hendry (2011).

We now illustrate the inaccuracy of some previous flash estimates, and the frequency and magnitude of location shifts, using examples for UK GDP outturns. First, initial and later estimates often differ substantively, as seen in Chart 1.

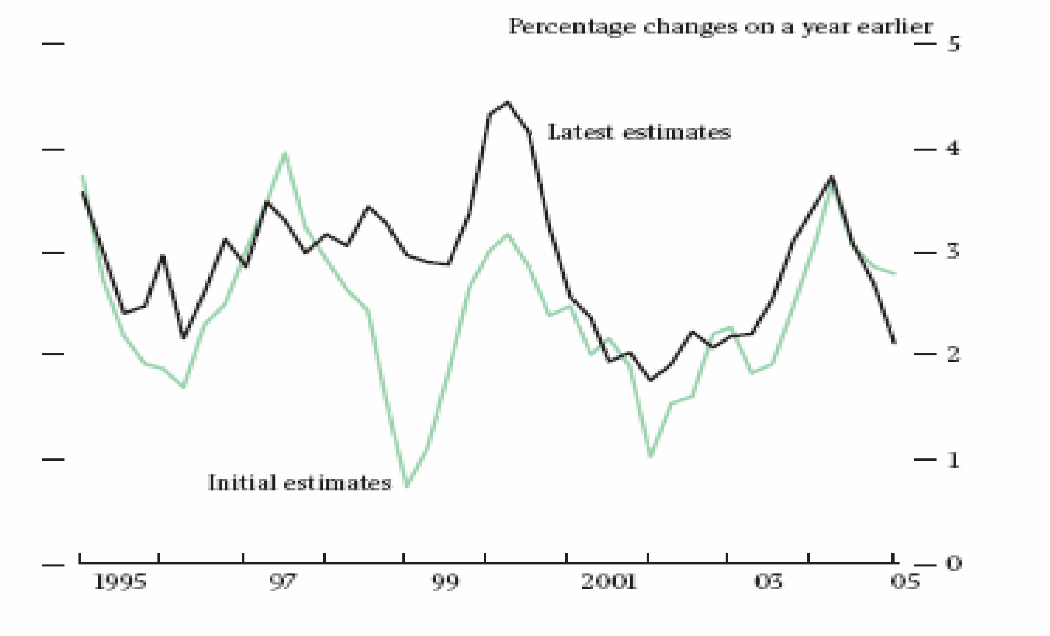


Chart 1. Source: Bank of England Quarterly Bulletin, Autumn 2005, p.338

Statistical agencies such as the UK Office for National Statistics (ONS) must base their flash estimates on incomplete and uncertain data. Hence, the need for nowcasts which combine both actual data for components that are known and forecasts of components that are unknown. Often surveys are used to 'infill' the unknown components to arrive at the flash estimate. This is a reasonable strategy when the surveys are good leading indicators, but such a methodology can prove disastrous in times of structural change. The *Financial Times* (25 July 2009) reported that 'the ONS said that its estimates were even less reliable than normal because the economy's unpredictability meant its models had "broken down".'[3] Any nowcasting strategy must be robust to structural breaks, requiring a methodology that can both detect structural breaks and rapidly adapt when such breaks occur. It is precisely at the point at which accurate nowcasts are most needed that current nowcasting methods are breaking down.

Secondly, unanticipated location shifts seem to occur all too often. As one of many possible examples, Barrell (2001) discusses six episodes of structural change in the 1990s (also see the cases documented by Stock and Watson, 1996, and Barrell, Dury, Holland, Pain and te Velde, 1998). Clements and Hendry (2001b) note the historical prevalence of failure in output forecasts for the UK, and the association of poor forecasts with major 'economic events'. Figure 1 records the recently measured changes in the log of seasonally-adjusted UK GDP at annual rates over 2008(1)–2011(2) highlighting the sharp drop during 2008.

The large empirical forecasting competitions, such as Makridakis, Andersen, Carbone, Fildes *et al.* (1982) and Makridakis and Hibon (2000), reviewed respectively by Fildes and Makridakis (1995) and Clements and Hendry (2001a), produce results across many models on numerous time series for several horizons. Their general findings are consistent with the implications of unanticipated location shifts being the major source of forecast failures, with other structural breaks being much less pernicious: see Hendry (2000). Thus, we consider the reasons why forecast failure occurs and the implications for how to produce robust forecasts in order to apply those lessons to nowcasting.

[3] www.ft.com/cms/s/0/c5529f00-7886-11de-bb06-00144feabdc0.html

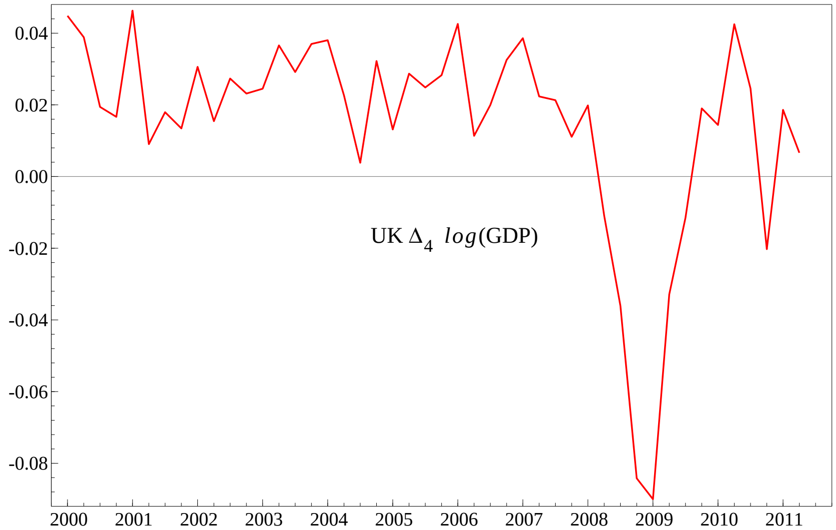


Figure 1: UK GDP year-on-year changes over 2000(1)–2011(2)

1.1 Chapter structure

The structure of the chapter is as follows. Section 2 surveys a number of extant nowcasting approaches as background. Section 3 then reviews some of the fundamental issues that affect both forecasting and nowcasting.

Section 4 reviews the theory of forecasting in the face of breaks, and seeks to explain forecast failure, and its absence, as many of the lessons learned there are germane to nowcasting. Readers familiar with that literature can go directly to section 5. Otherwise, in §4.1, we review the ten key difficulties confronting forecasting, using a scalar autoregressive-distributed lag (ADL) equation where the unmodeled variable is known into the future. Despite its simplicity, the illustration highlights all the main problems. §4.2 seeks to understand the resulting rather strange set of forecast errors, and shows that the unique explanation is the presence or absence of location shifts, namely shifts in the long-run mean of the variable being forecast. When such shifts occur, forecast failure results irrespective of the goodness or otherwise of the in-sample model; when they do not occur, there is no obvious failure irrespective of how many parameters of the data-generating process (DGP) shift, and by how much. §4.3 illustrates that robust forecasting devices may forecast better than any structural model in such shifting processes, as measured by root mean-square forecast errors (RMSFEs). §4.4 extends robust forecasting devices to first-order integrated processes (denoted I(1)). Section 4.5 provides three illustrative empirical applications of *ex post* forecasting, to Japanese exports over 2008(7)–2011(6), when they fell by more that 70% year on year, and UK GDP over two recessions.

The huge number of variables potentially involved raises the need for automatic model selection when there are more variables than observations, so section 5 describes the basis of automatic model selection, §5.1 outlines how *Autometrics* functions, §5.2 discusses the role of mis-specification testing, with the special but important case of impulse-indicator saturation (IIS) described in §5.3. Readers familiar with that literature can go directly to section 6, which considers using such findings to help guide the production of 'good' nowcasts.

All of these considerations apply to nowcasting an aggregate directly, as well as from disaggregates, which is the topic of §6. Then section 6.1 considers aggregation and forecasting aggregates from disaggregates given the ever-changing database of observed disaggregates, with subsections 6.2 on common features and the role factor models might play, §6.2.1 on the impacts of changing collinearity, §6.3 on using all available information including §6.3.1 on surveys, §6.3.2 on covariate information, including Google query data in §6.3.3, prediction market data in §6.3.4 and seasonal adjustment in §6.4. Next §6.5 considers specifying, selecting and estimating forecasting models for all disaggregates. Section 6.6 formulates the nowcasting strategy, where comparisons with early observed series suggests using any major discrepancies to adjust the forecasts of missing related disaggregates before calculating the preliminary estimates of the aggregate. §6.8 reviews the very practical problems of measurement errors and location shifts, and the role of judgment.

Section 7 applies the approach to *ex post* nowcasting UK GDP growth. §7.1 discusses the data, with §7.2 addressing the leading indicators considered. §7.3 explains our nowcasting methodology, §7.4 the actual specifications considered, with §7.5–§7.7 reporting the 'real-time' nowcasting results. Section 7.8 investigates the impacts and handling of contemporaneous breaks. Section 8 draws some conclusions.

2 Nowcasting approaches

In stable environments, delayed releases of preliminary aggregates may nevertheless still provide sufficiently timely and accurate indicators of the present economic state. For instance, the Bureau of Economic Analysis (BEA) produced preliminary estimates of quarterly US GDP that showed the correct direction of growth 98% of the time before 2008: see Fixler and Grimm (2008). However, in rapidly changing environments, preliminary releases may be both inaccurate and overdue, undergoing substantial revisions before the measurements are finalized, so can be poor approximations of the actual state of the economy.

The UK recession of 2008 provided an illustration of late and inaccurate reporting of dramatic macroeconomic developments. The first three preliminary estimates for the second quarter of 2008 produced by the ONS showed positive growth, when output had in fact dropped by 1.5%, a figure that was not reported until subsequent revisions. While national agencies may only get seriously wrong initial estimates infrequently, when they do, such mistakes can have important policy consequences. This underlines the need to produce alternative accurate real-time estimates of GDP growth, based on statistical nowcasting models using extensive information sets.

The majority of quantitative methodologies of nowcasting are based on projecting GDP growth onto a set of explanatory variables linked to aggregate output as in Bánbura *et al.* (2011). Relevant information sets often contain monthly or higher frequency indicators, involving both mixed-data frequencies and asynchronous releases, producing unbalanced panels with missing observations at the end of the sample, a problem often referred to as the 'ragged edge': see Wallis (1986).

Nowcasting can also be based on combining forecasts of disaggregated components. Lütkepohl (2006) provides a detailed survey of aggregate versus disaggregate forecasts, which is also applicable to nowcasting. When the DGP is known, disaggregated forecasts are at least as precise as a direct approach. When the DGP is unknown a-priori, Hubrich (2005) finds that direct forecasts of Eurozone HICP inflation have superior one-year-ahead predictive power to a combination of disaggregated forecasts, but this may not be the case when forecasting other variables or in different countries. Consequently, it is useful to understand the differences between direct nowcasts and a combination of disaggregated nowcasts as considered in Hendry and Hubrich (2011).

Notation can become overly complicated when every aspect of the changing environment of disaggregates and external variables and different frequencies are all denoted. In this section, we focus on frequency, so denote quarterly GDP by y_{t_q} where $t_q = 1, 2, \ldots, T_q$ represents quarters, with T_q representing the nowcast origin. GDP measures on a monthly frequency are denoted by $y_{t_m} = y_{t_q}$ and correspond to the months $t_m = 3t_q$, so that GDP is observed in month $t_m = 3, 6, 9, \ldots, T_m$, where $T_m = 3T_q$. Leading monthly indicators are denoted by z_{t_m} for $t_m = 1, 2, \ldots, T_m$, where the number of indicators is n. For each month t_m, the $t_m \times n$ matrix of exogenous explanatory variables is denoted by $\mathbf{Z}_{t_m}^1 = (\mathbf{z}_{t_m}, \ldots, \mathbf{z}_1)$, where $\mathbf{z}_{t_m} = (z_{1,t_m}, \ldots, z_{n,t_m})'$ is a vector of observations for a particular period. Given such information, the following nowcasting approaches are feasible.

Direct nowcasts

The least demanding approach is to nowcast the aggregate variable y_{t_q} using only its own past $\mathbf{Y}_{T_q-1}^1 = y_{T_q-1}, \ldots, y_1$:

$$\widehat{y}_{T_q|T_q-1} = f\left(\mathbf{Y}_{T_q-1}^1\right) \tag{1}$$

Specifying an autoregression for quarterly GDP estimates is a forecasting exercise using limited information of low frequency, so it is impossible to refine nowcasts between two instances of GDP publications. The latter becomes feasible if (1) is augmented by monthly indicators. Such a set of variables may include hard data like industrial production, financial indicators, interest rates measured at a higher frequency, or soft data like confidence or business surveys. Mitchell, Smith, Weale, Wright and Salazar (2005), Ferrara *et al.* (2010) and Bánbura *et al.* (2011) *inter alia* find that both types of indicators are relevant for explaining GDP growth.

In an augmented model, y_{t_q} is projected onto the set of monthly indicators $\mathbf{Z}_{t_m}$. A subset of $\mathbf{Z}_{t_m}$ is available for some vintage before T_q, say $T_q - v$, where $3T_m \leq T_q - v \leq 3\left(T_m + 1\right)$ and v is the maximum number of months for which the explanatory variable is available earlier than GDP. Since monthly indicators are released asynchronously with varying lags, a vector $(z_{1,T_m}, \ldots, z_{n,T_m})'$ will have zero entries for the unreleased variables. Consequently $\mathbf{Z}_{T_m}^1$ has ragged edges towards the end of the sample. Predictors are also released before the first GDP estimate is available. Therefore, the projection of y_{T_q} is conditioned on all available information and not just that contemporaneous to T_q. The expanding flow of new monthly releases means that v converges to 0 as time goes by, and so at each instance $T_q - v$, a new augmented model can be formulated conditional upon all available information:

$$\widehat{y}_{T_q|T_q-v} = f\left(\mathbf{Y}_{T_q-1}^1, \mathbf{Z}_{T_q-v}^1\right) \tag{2}$$

Indirect nowcasts using disaggregates information

The aggregate y_{t_q} is defined as a weighted sum of N_T disaggregated components y_{i,t_q} as $y_{t_q} = \sum_{i=1}^{N_T} w_i y_{i,t_q}$, where we ignore changes in weights. It is therefore possible to nowcast GDP as a weighted sum of individual component nowcasts. Assume that J_T out of N_T variables become available contemporaneously at time T_q, while $N_T - J_T$ disaggregates are only known up until lag $T - v^*$, where $v^* \geq v$, so not later than some of the indicators. The unknown disaggregates for $J_T + 1, \ldots, N_T$ are then nowcast using all available information at a disaggregated level:

$$\widetilde{y}_{i,T_q|T_q-v} = f\left(\mathbf{y}_{i,T_q-v^*}^1, \mathbf{Z}_{T_q-v}^1\right) \tag{3}$$

where $\widetilde{y}_{i,T_q|T_q-v}$ is the conditional nowcast and $\mathbf{y}_{i,t_q}^1$ are past observations for the component i. A combination of known and nowcast disaggregates yields the prediction of the aggregate:

$$\widetilde{y}_{T_q|T_q} = \sum_{i=1}^{J_T} w_i y_{i,T_q} + \sum_{i=J_T+1}^{N_T} w_i \widetilde{y}_{i,T|T_q-v} \tag{4}$$

Note that the aggregate nowcast $\widetilde{y}_{T_q|T_q}$ is contingent on a set of known components being released, which typically happens at the same time as the publication of the aggregate estimate. Consequently, nowcasting from (4) is possible at T_q, unless all components are unknown, needing nowcast at $T_q - v$.

Nowcasts using aggregated and disaggregated information

Two possibilities are available. To avoid the ragged edge problem, nowcasts are produced from a balanced panel of both aggregated and disaggregated variables. A direct nowcast is formulated at $T - v^*$, when the last disaggregate is observed:

$$\bar{y}_{T_q|T_q-v^*} = f\left(\mathbf{Y}^1_{T_q-v^*}, \mathbf{y}^1_{i,T_q-v^*}, \mathbf{Z}^1_{T_q-v^*}\right) \tag{5}$$

This approach is a simplification since it disregards discrepancies in the publication times between aggregate and disaggregate information. This may lead to exclusion of important dynamics for the variables available at a later period, but not yet included in the information set. Alternatively, it is possible to perform a quasi-in-sample nowcast by utilizing all available data from an unbalanced panel:

$$\bar{\bar{y}}_{T_q|T_q} = f\left(\mathbf{Y}^1_{T_q-1}; \mathbf{y}^1_{1,T_q}, \ldots, \mathbf{y}^1_{J_T,T_q}, \mathbf{y}^1_{J_T+1,T_q-v^*}, \ldots, \mathbf{y}^1_{N_T,T_q-v^*}; \mathbf{Z}^1_{T_q-v}\right) \tag{6}$$

Specification (6) is a generalization of the representation with ragged edges where the aggregate, disaggregate and leading indicators information is included and uses the most complete data set. A test of the four approaches is conducted in Hendry and Hubrich (2011), who compare direct aggregate forecasting with a combination of disaggregate forecasts using a Monte Carlo simulation as well as an empirical study of the US consumer price index, while conditioning on both types of information. They conclude that direct forecasts are more accurate than combinations of disaggregate forecasts. It is also found that adding interdependent disaggregated variables with different stochastic structures improves nowcasts of the aggregate for both approaches when a parsimonious specification is estimated.

Most nowcasting methodologies utilize all types of available information when nowcasting, and try to find ways of dealing with the ragged edge problem, rather than simplifying the analysis to balanced panels.

2.1 The 'in-filling' approach

Clements and Hendry (2003) reviewed the methodology used for nowcasting at the UK Office for National Statistics (ONS). The ONS then viewed forecasting any missing disaggregate data, or the aggregate directly, as a last resort for 'plugging gaps' when measured data were unavailable. 'In-filling' of missing disaggregates was based on either a version of the Holt–Winters exponentially-weighted moving average (EWMA: see Holt, 1957, and Winters, 1960) or an autoregressive-integrated moving average (ARIMA) model (see Box and Jenkins, 1976). The ONS guidelines as to which to use emphasized whether the series was sufficiently important to merit going through an ARIMA model-fitting exercise, and whether time and expertise were available to do so, and the implementation of the chosen method was reviewed annually.

Both EWMA and ARIMA are well designed for handling first-order integrated (I(1)) data generated by a random walk subject to measurement errors. However, that does not seem to be an optimal assumption about the process generating the disaggregates in an economy subject to intermittent and often unanticipated location shifts, as neither method is robust after such breaks. Moreover, knowing that shifts have occurred in some of the measured variables suggests that a similar problem may affect related unobserved series. Multivariate approaches, 'outside' explanatory variables (other than as proxies), and mixed-frequency data (see e.g., Montgomery, Zarnowitz, Tsay and Tiao, 1998) were not used to obtain more accurate nowcasts. Thus, improvements to in-filling seem feasible.

2.2 Mixed-data sampling models

Introduced in Ghysels, Santa-Clara and Valkanov (2004), the mixed-data sampling (MIDAS) approach links a low-frequency variable with selected predictors at higher frequency using a parsimonious restricted lag polynomial. Ghysels, Sinko and Valkanov (2007) and Andreou, Ghysels and Kourtellos (2011) provide comprehensive reviews of restrictions. The most common specification with coefficients of the lagged predictors being modeled as distributed lag functions is presented below. For simplicity of exposition, we consider a single indicator $z_{t_m} \in \mathbf{Z}_{t_m}$ explanatory variable. A MIDAS regression for a quarterly nowcasting horizon of $h_q = 3h_m$ quarters ahead, where h_m denotes the corresponding monthly horizon, is then given by:

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \beta_1 B\left(L_m; \boldsymbol{\theta}\right) z^{(3)}_{t_m+v} + \epsilon_{t_m+h_m} \tag{7}$$

where $B\left(L^m; \boldsymbol{\theta}\right)$ is the exponential Almon lag function with:

$$B\left(L^m; \boldsymbol{\theta}\right) = \sum_{k=0}^{K} B\left(k; \boldsymbol{\theta}\right) L^k_m, \qquad B(k; \boldsymbol{\theta}) = \frac{exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=0}^{K} exp(\theta_1 k + \theta_2 k^2)} \tag{8}$$

and where $L^k_m z_{t_m} = z_{t_m - k}$ is the monthly lag operator, and $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$ is the parameter vector. The superscript (3) in $z^{(3)}_{t_m+v}$ denotes that the predictors are observed with frequency three times as high as the dependent variable in the same time frame. Finally, K is the largest lag of the explanatory variable allowed to enter the regression. Since β_1 captures the overall effect of $z^{(3)}_{t_m+v}$ and its lags via $y_{t_m+h_m}$, identification of the coefficient is ensured by normalization: $\sum_{k=0}^{K} B\left(k; \boldsymbol{\theta}\right) = 1$. Given the estimated parameters $\widehat{\beta}_0, \widehat{\beta}_1$ and $\widehat{\boldsymbol{\theta}}$, the nowcast at $T_m + v$ is computed as:

$$y_{T_q+h_q|T_m+v} = \widehat{\beta}_0 + \widehat{\beta} B\left(L_m; \widehat{\boldsymbol{\theta}}\right) z^{(3)}_{T_m+v} \tag{9}$$

An extension of this basic model, proposed in Ghysels *et al.* (2004), considers the autoregressive representation:

$$y_{t_q+h_q} = \beta_0 + \lambda y_{t_m} + \beta_1 B\left(L_m; \boldsymbol{\theta}\right)\left(1 - \lambda L^3_m\right) z^{(3)}_{t_m+v} + \epsilon_{t_m+h_m} \tag{10}$$

where the autoregressive coefficient λ is estimated jointly with the lag polynomial parameters and the regression coefficients using non-linear least squares. An empirical implementation of an autoregressive MIDAS model is presented in Clements and Galvão (2008), who use a small subset of indicators that may be relevant in explaining GDP growth. They find that industrial production and capacity utilization significantly reduce the root mean-square error (RMSE) for US GDP nowcasts as compared to an autoregressive benchmark, similar to that in (1). In a subsequent application of the MIDAS approach, Clements and Galvão (2009a) find that single indicator models are not robust predictors of recessions, but multiple-indicators MIDAS models improve nowcasting accuracy relative to a forecast combination, especially when predicting the direction of growth. This suggests using extensive information to improve nowcasting performance, but estimation of multiple-variable MIDAS models is susceptible to mis-specification. The next model attempts to resolve that issue by summarizing all variation in the explanatory variables using a smaller number of combinations.

2.3 Factor models

The so-called 'curse of dimensionality' seems relevant in nowcasting since the information set may contain a large number of explanatory variables. Estimating all parameters may be impossible (an issue addressed in §5.3 below), whereas omitting relevant variables will result in mis-specification and reduced explanatory power. One solution to both these problems is to summarize the dynamics of the monthly indicators using a small set of common factors and apply those as explanatory variables. Boivin and Ng (2005) provide a summary of factor models for single frequency data. A general approach includes two steps. Initially the factors are computed, followed by an estimation of a regression for the dependent variable augmented with these factors.

A dynamic factor representation of a vector $\mathbf{x}_t$ has the following form:

$$x_{i,t} = \boldsymbol{\lambda}_i(L)'\mathbf{f}_t + \xi_{i,t} \tag{11}$$

where $\mathbf{f}_t$ is an $\bar{r} \times 1$ vector of dynamic common factors, $\boldsymbol{\lambda}_i(L)$ is a lag polynomial of degree s, L is the lag operator, and $\boldsymbol{\xi}_t$ is a $n \times 1$ vector of idiosyncratic components. Given this representation, Stock and Watson (1998) propose to stack the factors in an $r \times 1$ vector $\mathbf{F}_t = \left(\mathbf{f}'_t, \ldots, \mathbf{f}'_{t-s}\right)'$ where $r \leq \bar{r}(s+1)$ to obtain a static representation of a factor model:

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \tag{12}$$

where $\boldsymbol{\Lambda}$ is a matrix of loadings, with the ith row of $(\lambda_{i,0}, \ldots, \lambda_{i,s})$, of the common static factors $\mathbf{F}_t$. In both representations the idiosyncratic component is orthogonal to the factors. The common components are computed as:

$$\boldsymbol{\chi}_t = \boldsymbol{\Lambda}\mathbf{F}_t \tag{13}$$

In the classical case with no ragged edges, a balanced panel $\mathbf{Z}^1_{T_m}$ does not have have any missing observations at the end of the sample. In this case, there are two main estimation procedures considered in the literature. A classical approach proposed by Stock and Watson (2002a, 2002b) uses static principal components (PCs) of the sample covariance matrix. Principal components $\widehat{\mathbf{F}}$ of a $T \times n$ matrix of the explanatory variables are calculated as $\widehat{\mathbf{F}} = \mathbf{X}\widehat{\mathbf{H}}$ where $\widehat{\mathbf{H}} = \left(\widehat{\mathbf{h}}_1, \ldots, \widehat{\mathbf{h}}_n\right)$ is the matrix of eigenvectors corresponding to the largest eigenvalues $\left(\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_n\right)$, of the covariance matrix $\widehat{\mathbf{C}}$ of $\mathbf{X}$, given by the eigenvalue decomposition:

$$\widehat{\mathbf{C}} = T^{-1}\mathbf{X}'\mathbf{X} = \widehat{\mathbf{H}}\widehat{\boldsymbol{\Lambda}}\widehat{\mathbf{H}}'.$$

Dimensionality reduction is obtained by selecting a few factors corresponding to the largest eigenvalues that explain most of the variation in the data on $\mathbf{X}$.

In the second stage, an OLS projection of the dependent variable onto a set spanned by the factors $\widehat{\mathbf{F}}$ is estimated. As an alternative to static PCs, Forni, Hallin, Lippi and Reichlin (2001) consider an approach of estimating dynamic factors using generalized PCs where the weight of each observation is proportional to its signal-to-noise ratio. They use non-parametric techniques to forecast the factors taking into consideration restrictions on the dynamic factor structure. Forni, Hallin, Lippi and Reichlin (2005) argue that this method provides theoretical efficiency improvements as compared to a static factors representation.

Since the introduction of the static factor model, there have been various empirical studies confirming its superiority to models comprising a small number of variables: see Stock and Watson (2002b), Camacho and Sancho (2003), Brisson, Campbell and Galbraith (2003) and Artis, Banerjee and Marcellino (2005) inter alia. However, its performance has also been questioned in Giacomini and White (2006), who show that in a moving-window framework, static factors do not necessarily provide a forecasting accuracy improvement. Moreover, Banerjee, Marcellino and Masten (2058) for the Euro-area inflation and GDP, and Schumacher and Breitung (2008) for German data, show that static factors need not perform better than single indicator models. Castle, Clements and Hendry (2012) apply the approach in Castle, Doornik and Hendry (2012a) (who model using all the variables jointly with their PCs) to forecasting US GNP, but find that when the sample is extended over the recent recession, neither variables nor PCs (nor both) perform well.

Thus, despite various empirical and simulation based comparisons, there is no consensus on whether static or dynamic factor models forecast better. Forni, Hallin, Lippi and Reichlin (2003), who analyze financial time series linked to Euro-area inflation and GDP, D'Agostino and Giannone (2012) using a large cross-sectional data study for the US, and a survey in Stock and Watson (2006), all conclude that forecasting accuracy of the two specifications is similar so that it is difficult to distinguish a clear-cut winner. In particular, for the original US data set used in Stock and Watson (2002a), only a few factors provide substantial additional predictive power as compared to a simple autoregressive specification for forecasting of inflation and industrial production, and dynamic factors do not yield any improvement in predictive accuracy.

Conversely, supporting the empirical superiority of dynamic factors are the studies by den Reijer (2005) for the Dutch economy, Schumacher and Breitung (2007) for German data, and D'Agostino, McQuinn and O'Brien (2011) for Irish GDP, who find substantial accuracy improvement when using the approach in Forni *et al.* (2001). In the opposite camp is the evidence provided by Boivin and Ng (2005), who find that the static approach is robust to model mis-specification and performs better when forecasting variables with unknown dynamic structure, since it imposes no restrictions on the factors and forecasts the series directly. Overall, both approaches tend to achieve accuracy improvement as compared with benchmark models. Since there is no clear leader in this race, and much room for further investigation as data properties change over time, the static factor approach is slightly preferred for its practical simplicity.

2.4 Factors with ragged edges

In the context of nowcasting, factor estimation is more problematic in view of the missing observations at the end of the sample. The ragged-edge problem makes the panel unbalanced, so prevents direct estimation of static or dynamic factors using the methods just discussed. Therefore a procedure that can handle missing data is required.

A simple approach is discussed in Altissimo, Cristadoro, Forni, Lippi and Veronese (2010), who propose to use vertical re-alignment of the data. In practice, this implies that the most recent available data point is assigned to be the current value of the series. For instance, suppose that at time T_q some indicator z_i is available with a lag of l_i, so the most recently released value is z_{i,T_q-l_i}. Vertical re-alignment assigns this value to z_i at period T_q:

$$\widetilde{z}_{i,T_q} = z_{i,T_q-l_i} \tag{14}$$

When applied to all series for $t_l = l_i + 1, \ldots, T_q$, given that l_i is the maximum release lag, this procedures, results in a balanced panel $\widetilde{\mathbf{Z}}^1_{T_q}$. Despite being undemanding computationally, and for this reason convenient practically, this method distorts the dynamic correlation structure each time realignment is applied, since the data are released at different times and various revisions are applied: see Marcellino and Schumacher (2010).

An alternative way of constructing a balanced panel from data with missing observations is to use an EM algorithm for the static factors in (12), as suggested by Stock and Watson (2002b). The algorithm constructs a balanced panel through a recursive procedure that updates the missing observations and re-estimates the factors at the last column of the data matrix that contains zeros for the unobserved values. Angelini, Henry and Marcellino (2006) point out that although there are more sophisticated methods of data interpolation and backdating, such as large factor state-space representations, as in Kapetanios and Marcellino (2010), or utilizing the Kalman Filter as in Doz, Giannone and Reichlin (2011), the EM algorithm performs interpolation and estimation of the factors consistently and is computationally convenient. Given that many of the most recent studies, including D'Agostino and Giannone (2012), Schumacher and Breitung (2007, 2008) and Marcellino and Schumacher (2010) find that when correcting for the ragged-edges in nowcasting, the EM algorithm does not perform worse than the more computationally demanding methods, it may be preferred for its practical convenience.

In an empirical study of factor models, Bánbura *et al.* (2011) apply dynamic factors to nowcasting Euro-area GDP, and find that as new high-frequency information arrives throughout the reference quarter, nowcasting accuracy improves monotonically. Less timely hard data variables become important closer to the end of the quarter, while survey data provides an early indication for GDP growth. This is confirmed in Giannone, Reichlin and Simonelli (2009) and Bańbura and Rünstler (2011) who find that survey data provides a good early estimate of GDP, with real activity and financial data variables playing a relatively less important role. The same conclusion is reached by Giannone *et al.* (2008), who use dynamic factor models with the Kalman smoother updates for missing observations in a pseudo real-time US GDP nowcasting exercise. These findings provide additional justification for utilizing mixed frequency ragged-edged data of varying types in nowcasting as opposed to formulating balanced quarterly forecasting models.

The main difficulties for factor-based nowcasts are the ragged-edge problem and mixed data frequencies. As an alternative to using factors directly as explanatory variables, a hybrid MIDAS-factor model could be formulated. This approach combines all available information in a parsimonious way, thereby resolving the problems of mixed frequencies, ragged edges and dimensionality. In an empirical study, Marcellino and Schumacher (2010) and Kuzin, Marcellino and Schumacher (2009a) use a MIDAS-factor approach to nowcast GDP growth in six countries. For $r > 1$, they estimate common factors $\widehat{\mathbf{F}}_{t_m} = \left(\mathbf{f}'_{1,t_m}, \ldots, \mathbf{f}'_{r,t_m}\right)$ and use them in the following MIDAS regression:

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \sum_{i=1}^{r} \beta_{1,i} B_i \left(L_m; \theta_i\right) \widehat{\mathbf{F}}_{t_m}^{(3)} + \epsilon_{t_m+h_m} \tag{15}$$

where $B_i \left(L_m; \theta_i\right)$ are the Almon lag polynomials as in (8). Kuzin *et al.* (2009a) provide a comparison of the ragged edge correcting technique withing the MIDAS and MIDAS-factor frameworks. They find that all specifications have similar nowcasting performance, and that using real-time ragged edge information for nowcasting outperforms balanced panels. However, their results suggest that pooling single-indicator models tends to outperform factor models. Eickmeier and Ziegler (2008) reach the same conclusion based on US and UK data. This effect is robust to the selection of the pooling weighting schemes. Kuzin *et al.* (2009a) also conduct sequential model selection, and conclude that the results do not match the accuracy achieved by pooling. Their selection procedure is, however, based on a single simple criterion, and more sophisticated model selection techniques could well improve on this benchmark case. Another alternative to factor models and MIDAS regressions that assures dimension reduction and parsimonious model representation is provided by Bayesian methods. However, following De Mol, Giannone and Reichlin (2008) who compare Bayesian shrinkage techniques with factor models and find that the two are on par, discussion of the former is omitted here: a simpler and intuitively convenient nowcasting approach is described in the next sub-section.

2.5 Bridge equations

MIDAS regressions and factor models are parsimonious but restrictive ways of combining extensive information set with mixed frequency data and ragged edges. Moreover, their results are hard to interpret: the lag distribution of MIDAS coefficients is ad-hoc, and common factors may not be uniquely identifiable. PCs are a transformation of the original data, which provides no additional information about the underlying link between the target series and exogenous variables. For these reasons, a slightly less restrictive approach is based on equations that construct a direct 'bridge' between aggregate measures and a set of explanatory variables. Bridge equations are popular in central banks, mainly due to their interpretability. The approach specifies an equation for GDP growth, Δy_{t_q}, at a quarterly frequency, as in Mitchell (2009):

$$\Delta y_{t_q} = c + \sum_{i=1}^{p} \alpha_i \Delta y_{t_q - i} + \sum_{j=0}^{p} \sum_{i=1}^{k} \beta_{ij} z_{i,t_q - j} + u_{t_q} \text{ for } t = 1, \ldots, T \tag{16}$$

where k quarterly indicators z_{i,t_q} are transformed to stationarity, p is the lag length and u_t are assumed to be IID residuals. The regression parameters describe the link between the leading indicators and GDP growth, a feature missing in both the MIDAS and factor models.

The quarterly indicators of z_{i,t_q} in (16) are transformed versions of the observed monthly indicators, a subset of which is unobserved in-sample due to the ragged edge problem. The individual equations that forecast the missing values of the monthly indicators are then specified to correct for the unbalanced panels. In the simplest setting, a univariate AR(p) process could be used:

$$x_{i,t_m} = \sum_{i=1}^{p} \gamma_i x_{i,t_m - i} + e_{i,t_m} \tag{17}$$

where e_{i,t_m} is assumed to be white noise. The quarterly indicators are then formed using a combination of the observed monthly data, where available, and their forecasted values. Essentially, bridge equations provide an alternative tool for correcting for the ragged-edge problem by directly predicting the missing monthly observations required for nowcasting as in Mariano and Murasawa (2003). Also, Barhoumi, Benk, Cristadoro, Reijer, Jakaitiene, Jelonek, Rua, Ruth, Van Nieuwenhuyze and Rünstler (2008) find that utilizing monthly information to forecast the missing data tends to improve nowcasts compared with balanced quarterly regressions. Bridge equations can also take the form of a factor model, as described in Gertler and Rogoff (2005) and Giannone *et al.* (2008), such that:

$$\Delta y_{t_q} = c + \boldsymbol{\delta}' \mathbf{f}_{t_q} + u_{t_q} \tag{18}$$

where $\mathbf{f}_{t_q}$ are quarterly factors aggregated over their monthly representations $\mathbf{f}_{t_m}$ that are estimated from monthly data on the leading indicators using, for example, a dynamic representation:

$$x_{t_m} = \boldsymbol{\lambda}' \mathbf{f}_{t_m} + \xi_{t_m}$$

as in (12), and the missing factors are forecast using the Kalman Filter. In an empirical application of bridge equations, Mitchell (2009) considers regression-based nowcasts for UK GDP with the information set containing industrial production indicators and qualitative survey data from the Confederation of British Industry (CBI). A simple factor model is estimated, linking the factors obtained using bridge equations, that can successfully nowcast the start of the recession in 2008 two months before the publication of the official data by the ONS. In a similar fashion, Kitchen and Monaco (2003) and Rünstler and Sédillot (2003) apply bridge equations to US and Euro-area data respectively, and show that nowcasting improves on naive autoregressive and random-walk models.

Diron (2008) uses hard data, surveys and sentiment indicators to nowcast Euro-area GDP growth, and shows that nowcasts based on pseudo real-time forecasts of monthly indicators using bridge equations differ to nowcasts obtained using revised data. However, revisions to GDP and monthly indicators constitute only a small fraction of the nowcasting error. This may not be the case for the UK, since the ONS failed to measure a decline in output in the second quarter of 2008 at the time it actually happened.

Given that most studies concentrate on relatively stable periods with no major macroeconomic perturbations, the finding in Diron (2008) provides reassurance that nowcasting methodologies can improve not only on simple benchmark models, but also perform better than the measurements by the statistical agencies to yield more timely and accurate estimates. This is confirmed by the ex-post nowcasting exercise in Mitchell (2009) who essentially predicted the recession in the UK half a year before the ONS reported a decline in output. Moreover, the ONS estimate was published only after an additional quarter had elapsed. These findings suggest that bridge equations provide an easy, interpretable framework for nowcasting that performs as well as more sophisticated and computationally demanding models. Raw implementation of bridge equations in large information sets is, however, subject to a dimensionality curse, since the number of parameters to be estimated grows linearly with the number of monthly indicators n. A feasible solution to that problem is to use model selection, as discussed in §5.

We now consider the difficult issue of how location shifts affect general model formulation.

3 Some fundamentals

In this section, we show the key result that when unanticipated location shifts occur, conditional expectations of a future variable need not be unbiased even based on the in-sample DGP and given all available information. Thus, other devices for nowcasting and forecasting may outperform.

The theory of economic forecasting when an econometric model coincides with a stationary DGP has been well developed along lines first proposed in Haavelmo (1944). Consider an $n \times 1$ vector $\mathbf{x}_t \sim \mathsf{D}_{\mathsf{x}_t}(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta})$ for parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathcal{R}^k$, where $\mathsf{D}_{\mathsf{x}_t}(\cdot)$ denotes the data density of $\mathbf{x}_t$, conditional on $\mathbf{X}_{t-1} = (\ldots \mathbf{x}_1 \ldots \mathbf{x}_{t-1})$. A forecast $\widetilde{\mathbf{x}}_{T+h|T} = \mathbf{f}_h(\mathbf{X}_T)$ is desired at T for $\mathbf{x}_{T+h}$ in time $T+h$ using all the available information. The main issue is how to select $\mathbf{f}_h(\cdot)$, and the well-known answer is the conditional expectation, $\widehat{\mathbf{x}}_{T+h|T} = \mathsf{E}[\mathbf{x}_{T+h}|\mathbf{X}_T]$ which is unbiased:

$$\mathsf{E}\left[\left(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right) \mid \mathbf{X}_T\right] = \mathsf{E}[\mathbf{x}_{T+h}|\mathbf{X}_T] - \mathsf{E}[\mathbf{x}_{T+h}|\mathbf{X}_T] = \mathbf{0} \tag{19}$$

and $\widehat{\mathbf{x}}_{T+h|T}$ has the smallest mean-square forecast-error matrix of unbiased predictors:

$$\mathsf{M}\left[\widehat{\mathbf{x}}_{T+h|T} \mid \mathbf{X}_T\right] = \mathsf{E}\left[\left(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right)\left(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right)' \mid \mathbf{X}_T\right] \tag{20}$$

Such results sit uneasily with the intermittent occurrence of forecast failure based on models that seek to capture conditional expectations. Either econometric models fail badly in their aim, or the results in (19) and (20) are not empirically relevant. Since the DGP is not known, the former is quite possible: there are many problems learning about $\mathsf{D}_{\mathbf{X}_T^1}(\cdot)$ and $\boldsymbol{\theta}$, involving the specification of the set of relevant variables $\{\mathbf{x}_t\}$, measurement of those $\mathbf{x}$s, the formulation of $\mathsf{D}_{\mathbf{X}_T^1}(\cdot)$, modeling the relationships between the variables, and estimating $\boldsymbol{\theta}$, all of which introduce in-sample uncertainties. The unknown properties of $\mathsf{D}_{\mathbf{X}_{T+H}^{T+1}}(\cdot)$ determine the forecast uncertainty, which grows as H increases, especially for integrated data. Thus, the conventional answer for the failure of (19) and (20) to hold in practice is that we fail to model conditional expectations well enough.

However, results from our own research provide a framework for explaining why forecast failure can occur even when the best structural model is used, which coincides with the DGP in-sample, and indeed even when the future values of all unmodeled variables are known before forecasting. Unanticipated location shifts discussed in §1 have a profound implication: conditional expectations formed at time T from information available at that date for an outcome at $T+1$ are neither unbiased nor minimum variance when unanticipated location shifts occur (see Hendry and Mizon, 2010, 2011b). When the mean of the data distribution has shifted from T to $T+1$, the previous conditional expectation is not centered on the relevant new mean. Since many theorems about economic forecasting begin with the presumption that conditional expectations provide unbiased and minimum variance predictors, and hence deduce that econometricians should use the conditional expectations of models, a major rethink is required. Thus, unanticipated changes in $\mathsf{D}_{\mathbf{X}_{T+H}^{T+1}}(\cdot)$ and $\boldsymbol{\theta}$ are the culprit. It is manifest that similar implications hold for nowcasts.

Next, rather than being a pernicious necessity to be avoided at all possible costs, model selection by reduction in a general-to-specific (*Gets*) approach is usually beneficial. A number of studies have demonstrated that unless the prior theory is precisely correct, selection delivers improved models (see Doornik, 2009a, Doornik and Hendry, 2009, and Castle, Doornik and Hendry, 2011a, and Castle and Hendry, 2012). If the theory model is correct, then even after considerable search, selection can still deliver identical estimates for its parameters to those obtained by direct fitting (see Hendry and Johansen, 2012). By using an extended *Gets* approach, there can even be more candidate variables to be selected from than the number of observations (see Hendry and Krolzig, 2005, and Doornik, 2007). In turn that allows checking for possible location shifts at every observation to be handled during selection using impulse-indicator saturation (IIS: see Hendry, Johansen and Santos, 2008, and Johansen and Nielsen, 2009, for analyses). Castle, Doornik and Hendry (2012b) investigate the performance of IIS in detecting multiple location shifts during selection, including breaks close to the start and end of the sample (as well as correcting for non-normality), and Hendry and Mizon (2011a) provide an empirical application to demand for food in the USA.

Finally, and a consequence of the previous findings, the quality of forecasts from a model depends both on how it is used and what the properties of the forecast period are, as much as on the specification of the model itself and how it was estimated (see e.g., Clements and Hendry, 1998, 1999, and Hendry, 2006). We structure our analysis of nowcasting around these forecasting issues.

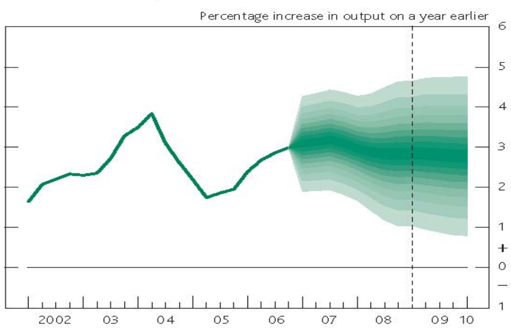
In part, the increasing realization of unanticipated location shifts as a key problem when modeling and forecasting has necessitated an emphasis on the invariants of the economic process, first discussed by Frisch (1938) and later highlighted by Lucas (1976), inducing a reformulation of the notion of exogeneity (see Engle, Hendry and Richard, 1983, and Engle and Hendry, 1993). However, the problem of unanticipated shifts was central to the prescient work of Smith (1927, 1929), which somehow was forgotten despite being published just before the Great Depression (see Mills, 2010, who focuses on Bradford Smith's research on the 'nonsense regressions' problem). Moreover, new tests of the key concepts of shifts and invariance have appeared, including several based on impulse-indicator saturation (see e.g., Hendry and Santos, 2010). Thus, computational power and hugely improved algorithms have built on theoretical understanding to radically alter how econometric models can be constructed, and should be used.

4 Explaining forecast failure–and its absence

In this section, we illustrate the key result that the presence or absence of location shifts is the main determinant of forecast failure, such that in open models with zero intercepts, even knowing the future values of unmodeled variables and using the in-sample DGP need not avoid forecast failure. Model mis-specification and parameter estimation uncertainty play secondary roles, illustrated by robust forecasting devices performing better than a structural model.

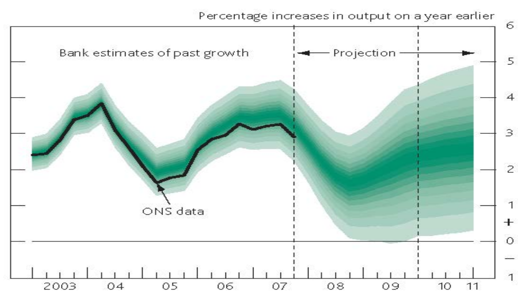
Forecasting is difficult. The next Chart from the *Bank of England Inflation Report* for February 2007 shows its forecasts at that time for annual changes in UK GDP: 'steady as she goes' through the end of 2009. The next Chart updates their forecasts to February 2008 through the end of 2010: a distinct slowdown is now envisaged. However, it is nothing like the 9% fall at annual rates that materialized, which was unanticipated at that stage: Figure 1 recorded the actual changes in UK GDP at annual rates over 2008(1)–2011(2). As seen, there was a sharp drop at the end of 2008. In late 2006 and very early 2007, there was little to suggest the UK's 16 years of uninterrupted growth was about to end: even if the US sub-prime mortgage problem had been high on the Bank's radar, it is unlikely anyone foresaw it nearly bringing the whole world economy to its knees.

Chart 1 Current GDP projection based on market interest rate expectations



The fan chart depicts the probability of various outcomes for GDP growth in the future. If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that GDP growth over the subsequent three years would lie within the darkest central band on only 10 of those occasions. The fan chart is constructed so that outturns of GDP growth are also expected to lie within each pair of the lighter green areas on 10 occasions. Consequently, GDP growth is expected to lie somewhere within the entire fan chart on 90 out of 100 occasions. The bands widen as the time horizon is extended, indicating the increasing uncertainty about outcomes. See the box on pages 48–49 of the May 2002 *Inflation Report* for a fuller description of the fan chart and what it represents. The dashed line is drawn at the two-year point.

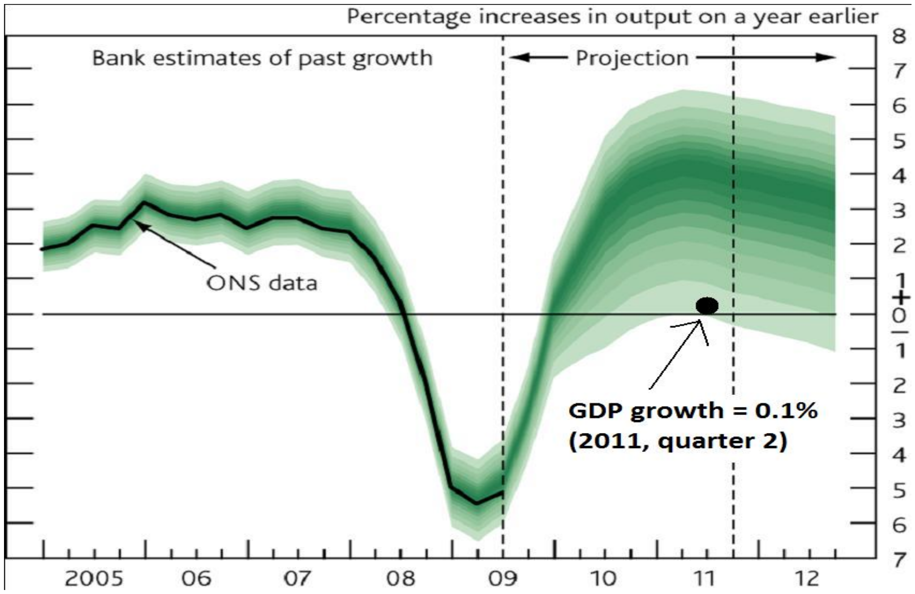
Chart 1 GDP projection based on market interest rate expectations



The fan chart depicts the probability of various outcomes for GDP growth. To the left of the first vertical dashed line, the distribution reflects the likelihood of revisions to the data over the past; to the right, it reflects uncertainty over the evolution of GDP growth in the future. If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that the mature estimate of GDP would lie within the darkest central band on only 10 of those occasions. The fan chart is constructed so that outturns are also expected to lie within each pair of the lighter green areas on ten occasions. Consequently, GDP growth is expected to lie somewhere within the entire fan on 90 out of 100 occasions. The bands widen as the time horizon is extended, indicating the increasing uncertainty about outcomes. See the box on page 39 of the November 2007 *Inflation Report* for a fuller description of the fan chart and what it represents. The second dashed line is drawn at the two-year point of the projection.

Source: Bank of England

Even a more recent set of Bank forecasts reveal the continuing difficulty of accurate forecasts as the following chart shows: the outcome in 2011(2) is at the lowest end of the range. Moreover, their uncertainty in the past outcomes is non-negligible.



Source: Bank of England

Such graphs emphasize that economies are non-stationary and evolving, where even the best economic model differs from the DGP, and forecast failures occur. Such failures are also associated with econometric model forecasts being out-performed by so-called 'naive devices', and that dates from the early history of econometrics, partly because almost no forecasting models allow for unanticipated location shifts, although these clearly occur empirically. Note that adding variables which 'explain' shifts in-sample will improve forecasts only if their shifts can be forecast in turn, but again that need not entail that nowcasts cannot be improved. We now focus on 1-step ahead forecasts as the aim is to elucidate corresponding issues for nowcasting.

4.1 Scalar autoregressive-distributed lag example

Consider an autoregressive-distributed lag DGP with a known exogenous $\{z_t\}$:[4]

$$x_t = \mu + \rho x_{t-1} + \gamma z_t + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{21}$$

which is stationary in sample with all parameters (ρ, γ, σ_ϵ^2) constant and $|\rho| < 1$. When ρ, γ are known and x_T and z_{T+1} are observed without error, the optimal forecast for x_{T+1} is:

$$\widehat{x}_{T+1|T} = \mu + \rho x_T + \gamma z_{T+1} \tag{22}$$

producing an unbiased forecast:

$$\mathsf{E}\left[\left(x_{T+1} - \widehat{x}_{T+1|T}\right) \mid x_T, z_{T+1}\right] = \mathsf{E}\left[\mu - \mu + (\rho - \rho)\, x_T + (\gamma - \gamma)\, z_{T+1} + \epsilon_{T+1}\right] = 0$$

which is zero with the smallest possible conditional and unconditional variance determined by $\mathsf{D}_{\mathbf{X}_T^1}(\cdot)$:

$$\mathsf{V}\left[\left(x_{T+1} - \widehat{x}_{T+1|T}\right) \mid x_T, z_{T+1}\right] = \mathsf{V}\left[\epsilon_{T+1}\right] = \sigma_\epsilon^2.$$

[4]This example is taken from Hendry and Mizon (2012a), which should be cited as the original source, as its use here is to provide a complete analysis. We include a contemporaneous 'exogenous' variable to link more closely with nowcasting.

Here, $\mathsf{D}_{\mathbf{X}_T^1}(\cdot)$ implies $\mathsf{D}_{\mathbf{X}_{T+1}^{T+1}}(\cdot)$ and $\mathsf{D}_{\mathbf{X}_{T+1}^{T+1}}(\cdot) = \mathsf{IN}\left[\mu + \rho x_T + \gamma z_{T+1}, \sigma_\epsilon^2\right]$. The focus of much of the literature on forecasting has been on the additional consequences of estimating the parameters of structural models like (21), often assumed to coincide with the DGP.

However, there is a long list of potential problems. These include that the specification may be incomplete if (e.g.) $\mathbf{x}_t$ is not a scalar; measurements might be incorrect if (e.g.) $\widetilde{\mathbf{x}}_t$ was observed, not $\mathbf{x}_t$; the formulation might be inadequate if (e.g.) an intercept was needed or z_t was omitted; the modeling might have gone wrong if (e.g.) x_{t-2} was incorrectly selected; estimating ρ adds a bias, $(\rho - \mathsf{E}[\widehat{\rho}])x_T$, and a variance $\mathsf{V}[\widehat{\rho}]x_T^2$ (as does estimating γ if z_t is included); the forecast analysis earlier in this section assumed $\epsilon_{T+1} \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$ but $\mathsf{V}[\epsilon_{T+1}]$ could differ; a multi-step forecast error $\sum_{h=1}^{H} \rho^{h-1}\epsilon_{T+h}$ would arise when $H > 1$, like that in the first Chart above, leading to $\mathsf{V}[\sum_{h=1}^{H} \rho^{h-1}\epsilon_{T+h}] = \frac{1-\rho^{2H}}{1-\rho^2}\sigma_\epsilon^2$ so if $\rho = 1$, the forecast variance would be trending as $H\sigma_\epsilon^2$; and if μ, ρ or γ changed, forecast failure could occur. Interactions between all of these difficulties could compound the problem for forecasters. Fortunately, as shown in Clements and Hendry (1999), most of the problems do not lead to the kind of mis-forecasting seen in the charts above, as we now illustrate. Unfortunately, as we also illustrate, even after determining the source of such forecast failures, many serious difficulties remain for the structural modeler.

To simplify the first round of analysis, we set $\mu = 0$ in the DGP (21) where this is known. That enables us to contrast forecasts like (22) knowing (ρ, γ) with estimating those parameters from a sample of accurate data over $t = 1, \ldots, T$:

$$\widetilde{x}_{T+1|T} = \widehat{\rho} x_T + \widehat{\gamma} z_{T+1} \tag{23}$$

Next, we mis-specify the model by omitting z_t, then also change ρ. The graphs that follow are based on a single random draw of $T = 50$ observations from (21) with $z_t \sim \mathsf{IN}[0, 1]$ when $\rho = 0.8$, $\gamma = 1$, and $\sigma_\epsilon^2 = 1$. A later theoretical analysis will conform the results are general despite the specificity of the illustration, but the graphs highlight the key issues.

Figure 2 panel a, records the forecasts both when ρ and γ are known and constant, and when they are estimated. $\widehat{x}_{T+h|T+h-1}$ from (22) is shown with error bars of $\pm 2\widehat{\sigma}$, whereas forecasts when estimating (ρ, γ) are shown as $\widetilde{x}_{T+h|T+h-1}$ with forecast interval bands. As can be seen, the two sets of forecasts are almost identical, with only a small increase in uncertainty from estimation. Thus, estimation *per se* does not seem to be a major problem, and is well known to be of probabilistic order $\mathsf{O}_p(T^{-1})$.

Next, we consider the impact on forecast accuracy and precision of incorrect specification, here corresponding to inadvertently omitting z_t, shown in Figure 2 panel b. When z_t is omitted both in estimation and forecasting the resulting forecasts are:

$$\widetilde{x}_{T+1|T} = \widetilde{\rho} x_T \tag{24}$$

These are clearly poorer than under correct specification, but well within the *ex ante* forecast intervals shown. Thus, forecast failure is not just a mis-specification problem either.

Finally in this group, at $T = 40$ we shift ρ to $\rho^* = 0.4$, then back to $\rho = 0.8$ at $T = 46$ till $T = 50$, so the DGP reverts to its previous state, mimicking a short, sharp regime switch in the dynamics, denoted:

$$x_{T+h} = \rho^* x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \tag{25}$$

over $h = 1, \ldots, 5$, before again becoming:

$$x_{T+h} = \rho x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \tag{26}$$

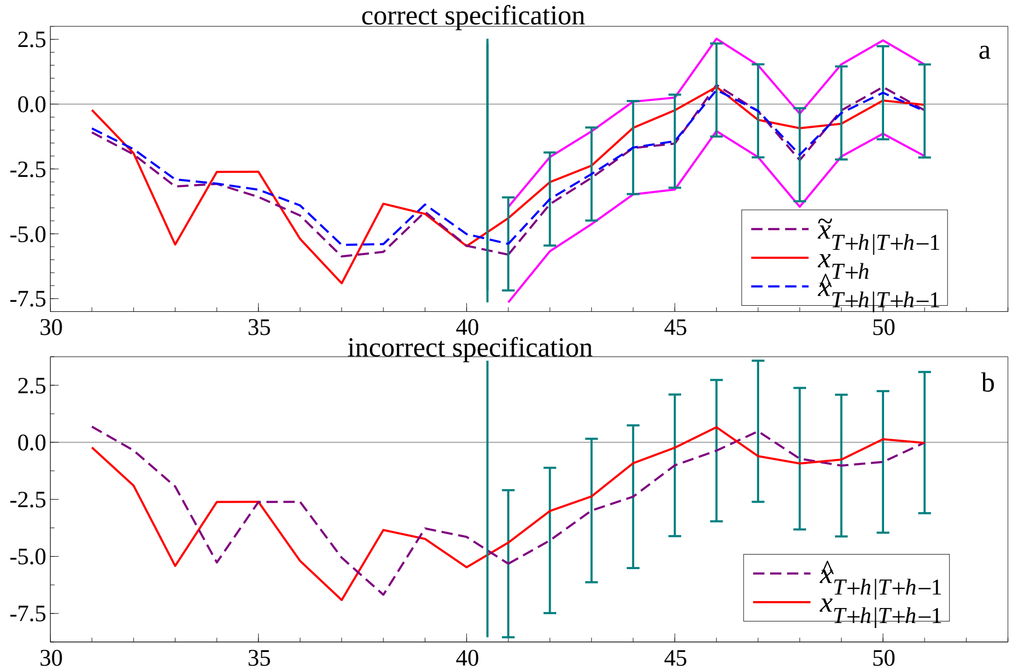


Figure 2: Forecasts under correct and incorrect specification

over $h = 6, \ldots, 10$. To allow for possible interactions with mis-specification, we first return to forecasting from (23), then also use (24). Figure 3 panel c reports the former: there is only a slight impact from halving ρ, then almost none from doubling it again, so forecast failure is not just a problem of changing parameters.

Next, we use (24), so all of estimation uncertainty, mis-specification, and non-constancy occur, yet again there is little noticeable impact from halving then doubling ρ as seen in figure 3 panel d. Whatever underlies outcomes as in the GDP Figure 1, those three mistakes do not explain it. Nor could reasonable measurement errors, nor a shift in γ, irrespective of the inclusion or omission of z_t (see Hendry and Mizon, 2012b), nor using the wrong lag value, such as x_{t-2}. Of course, none of these additional mistakes will help, but they are not the source of the problem of forecast failure.

Now reconsider the same shifts in ρ but where $\mu = 10$ in:

$$x_t = \mu + \rho x_{t-1} + \gamma z_t + \epsilon_t \tag{27}$$

Irrespective of the correct specification of the model in-sample–whether (23) or (24) is used–there is a catastrophic impact from halving ρ, here shown in figure 4 panel e for forecasts from (23). The data are not unlike those for UK GDP, first dropping then returning. Here, the first 5 forecasts are badly biased, and each is **above** the previous outcome despite the plunging values of x. Yet forecast failure vanishes on doubling ρ again. Manifestly, the value of μ, often treated as a nuisance parameter, is actually fundamental. So is the problem just the non-zero value of the intercept–or an interaction with model mis-specification?

To clarify, again for the model which was correctly specified in-sample, consider forecasts from (23) for the same breaks in ρ as in (25) and (26), but setting $\mu = 0$, where that is known to the investigator. Instead, we let $\mathsf{E}[z_t] = \kappa = 10$, noting that the model is correctly specified in-sample, there is a zero

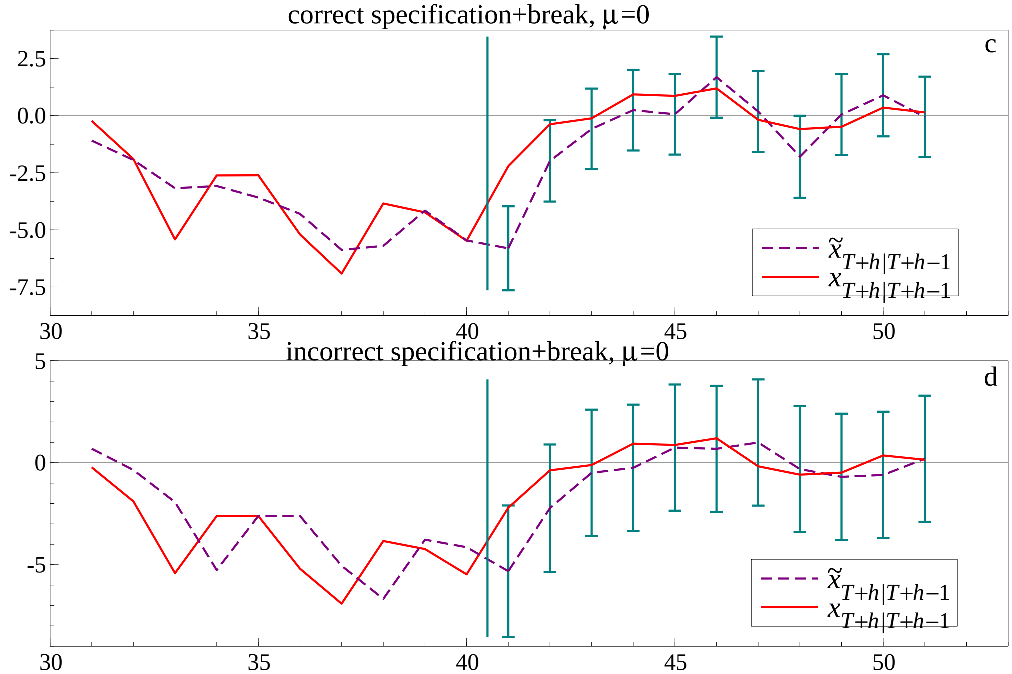


Figure 3: Correct and incorrect specifications with changed ρ

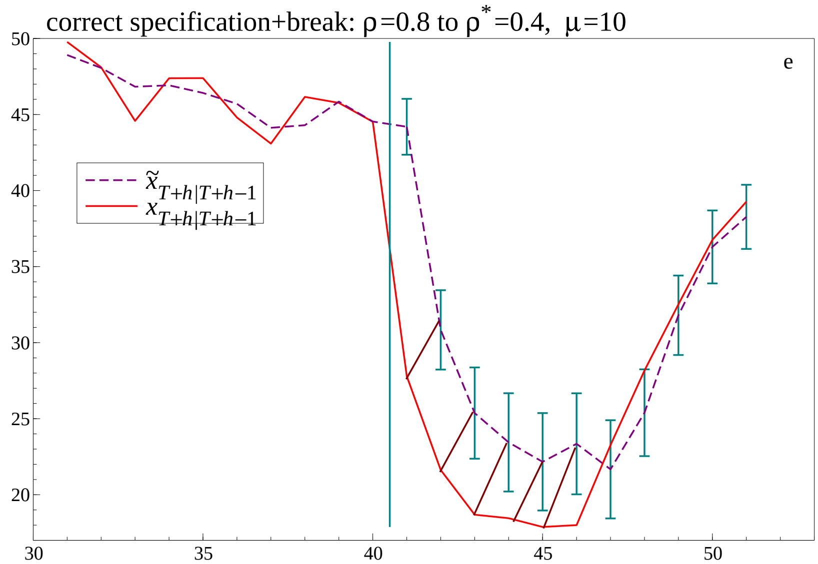


Figure 4: Correct in-sample specification with non-zero θ and changed ρ

intercept, and the forecasts use known future z_{T+h}. Despite all those advantagous features, forecast failure is again manifest in figure 4 panel f: it is almost identical to the outcome in panel e.

Even that is far from the whole story as can be seen as follows. When the model is incorrectly specified by omitting z_t as in (24), forecasting after the same breaks in ρ as in (25) and (26), but now with both $\mu = 10$, *and* $\kappa = 10$ where the former is also shifted hugely to $\mu^* = 50$ at $T = 41$ then back to $\mu = 10$ at $T = 46$ so:

$$x_{T+h} = \mu^* + \rho^* x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \tag{28}$$

Although almost everything seems to be wrong, there is no forecast failure using $\widehat{x}_{T+h|T+h-1} = \widehat{\mu} + \widehat{\rho} x_{T+h-1}$ over the ten forecasts in figure 5.

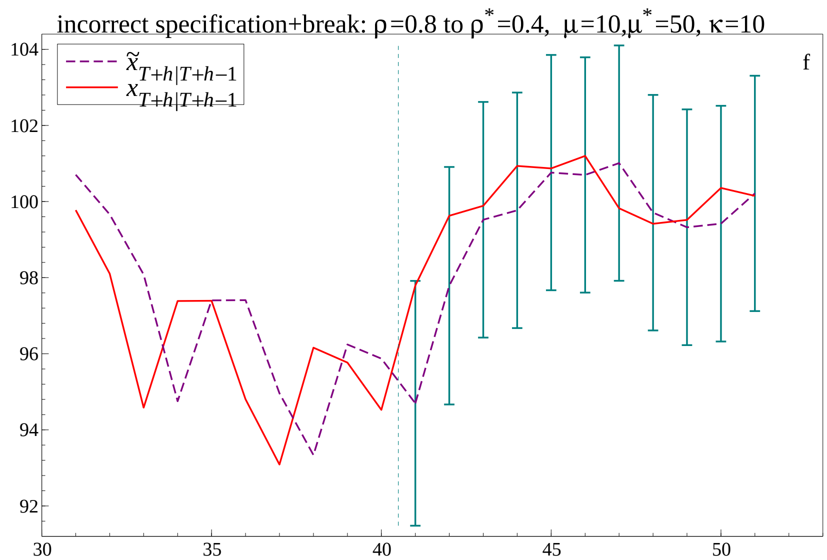


Figure 5: Incorrect specification with μ and ρ changed twice

4.2 Understanding these forecast errors

The unique explanation for this rather strange set of outcomes is in fact simple: location shifts, or alternatively expressed in this context, shifts in the long-run mean. When they occur, forecast failure results irrespective of the goodness or otherwise of the in-sample model; when they do not occur, there is no obvious failure irrespective of model mis-specification, how many DGP parameters shift and by much. Changes in $\mathsf{E}[x_t]$ are the culprit.

Let $\lambda = \gamma/(1-\rho)$ and rewrite the DGP in (21) as:

$$\Delta x_t = (\rho - 1)\left(x_{t-1} - \theta - \lambda\left(z_t - \kappa\right)\right) + \epsilon_t \tag{29}$$

and note that in general from (21) or (29):

$$\mathsf{E}[x_t] = \frac{(\mu + \gamma\kappa)}{(1-\rho)} = \theta \neq 0.$$

In the first three cases, and the last, $\mathsf{E}[x_{T+h}] = 0$ before and after the shifts in ρ. In the second set of cases, $\mathsf{E}[x_{T+h}]$ shifts from $\theta = 50$ to $\theta^* = 17$ in both cases *e* and *f* but $\theta = \theta^*$ in g. All models in this class are equilibrium correction so fail systematically when $\mathsf{E}[x_{T+h}]$ changes from θ to θ^*, as forecasts converge back to θ, irrespective of the new parameter values in the DGP. The class of equilibrium-correction models (EqCMS) is huge: all regressions; dynamic systems; VARs; DSGEs; ARCH; GARCH; and some other volatility models. Location shifts are a pervasive and pernicious problem affecting all EqCMS. To establish that claim formally, we next explain the taxonomy of all possible sources of forecast errors based on Hendry and Mizon (2012b).

The DGP is as in (21), but written as:

$$x_t = \theta + \rho\left(x_{t-1} - \theta\right) + \gamma\left(z_t - \kappa\right) + \epsilon_t \tag{30}$$

where $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$, $\mathsf{E}\left[x_t\right] = \theta$ and $\mathsf{E}[z_t] = \kappa$ with $\gamma \neq 0$, but z_t is omitted from the forecasting model:

$$x_t = \mu + \rho x_{t-1} + v_t$$

Here we only consider 1-step ahead forecasts facing a single break, which occurs immediately after time T, with the post-break DGP from $h = 1, \ldots$:

$$x_{T+h} = \theta^* + \rho^*\left(x_{T+h-1} - \theta^*\right) + \gamma^*\left(z_{T+h} - \kappa^*\right) + \epsilon_{T+h} \tag{31}$$

The estimated, mis-specified, forecasting model is denoted:

$$\widehat{x}_{T+1|T} = \widehat{\theta} + \widehat{\rho}\left(\widehat{x}_T - \widehat{\theta}\right) \tag{32}$$

estimated over $t = 1, \ldots, T$, with parameter estimates $(\widehat{\theta}, \widehat{\rho})$. Let $\mathsf{E}[\widehat{\theta}] = \theta_e$ and $\mathsf{E}[\widehat{\rho}] = \rho_e$. Forecasting takes place from an estimated $\widehat{x}_T$ at the forecast origin and yields the forecast error $\widehat{\epsilon}_{T+1|T} = x_{T+1} - \widehat{x}_{T+1|T}$:

$$\widehat{\epsilon}_{T+1|T} = \theta^* - \widehat{\theta} + \rho^*\left(x_T - \theta^*\right) - \widehat{\rho}\left(\widehat{x}_T - \widehat{\theta}\right) + \gamma^*\left(z_{T+1} - \kappa^*\right) + \epsilon_{T+1} \tag{33}$$

All the main sources of forecast error occur using (32) when (31) is the DGP:
stochastic breaks: $(\rho,\ \gamma)$ changing to $(\rho^*,\ \gamma^*)$;
deterministic breaks: $(\theta,\ \kappa)$ shifting to $(\theta^*,\ \kappa^*)$;
omitted variables: z_t excluded over both estimation and forecast periods;
biased (and inconsistent) parameter estimates: $\rho_e \neq \rho,\ \theta_e \neq \theta$ in general;
estimation uncertainty: $\mathsf{V}[\widehat{\rho},\ \widehat{\theta}] \neq \mathbf{0}$;
forecast-origin uncertainty: $\widehat{x}_T$;
innovation errors: ϵ_{T+h}.
The taxonomy of sources of forecast errors reveals all the possible effects, although we have omitted some small order interaction terms and estimation covariances of $\mathsf{O}_p(T^{-1})$ for simplicity. The calculations involved expand every term so that each final component corresponds to a single effect (e.g.):

$$\theta^* - \widehat{\theta} = \left(\theta^* - \theta\right) + \left(\theta - \theta_e\right) + \left(\theta_e - \widehat{\theta}\right) \tag{34}$$

so it is decomposed into shift, bias and estimation effects. Most previous taxonomies have focused on closed models, but reach similar conclusions to those discussed here.

Taxonomy for 1-step ahead forecast errors

$\widehat{\epsilon}_{T+1\mid T} \simeq$	Element	Expectation	Variance
$(1-\rho^*)(\theta^*-\theta)$	(*ia*)	$(1-\rho^*)(\theta^*-\theta)$	0
$+(\rho^*-\rho)(x_T-\theta)$	(*ib*)	0	$(\rho^*-\rho)^2\mathsf{V}[x_T]$
$+(1-\rho)(\theta-\theta_e)$	(*iia*)	$(1-\rho)(\theta-\theta_e)$	0
$+(\rho-\rho_e)(x_T-\theta)$	(*iib*)	0	$(\rho-\rho_e)^2\mathsf{V}[x_T]$
$-\rho(\widehat{x}_T-x_T)$	(*iii*)	$-\rho(\mathsf{E}[\widehat{x}_T]-x_T)$	$\rho^2\mathsf{V}[\widehat{x}_T-x_T]$
$-(1-\rho)\left(\widehat{\theta}-\theta_e\right)$	(*iva*)	0	$\mathsf{O}_p(T^{-1})$
$-(\widehat{\rho}-\rho_e)(x_T-\theta)$	(*ivb*)	$\simeq 0$	$\mathsf{O}_p(T^{-1})$
$+\gamma^*(z_{T+1}-\kappa^*)$	(*v*)	0	$(\gamma^*)^2\mathsf{V}[z_{T+1}]$
$+\epsilon_{T+1}$	(*vi*)	0	σ_ϵ^2

(35)

We now consider the implications of the forecast-error taxonomy, commencing from the foot of the table:
(*vi*): the innovation error has $\mathsf{E}[\epsilon_{T+1}]=0$ and $\mathsf{V}[\epsilon_{T+1}]=\sigma_\epsilon^2$ so there is no forecast bias, and an $\mathsf{O}_p(1)$ variance, which is irreducible if $\{\epsilon_t\}$ is an innovation;
(*v*): the omitted variable leads to $\mathsf{E}[\gamma^*(z_{T+1}-\kappa^*)]=0$ and $\mathsf{V}[\gamma^*(z_{T+1}-\kappa^*)]=\sigma_z^2$, so again induces no bias despite its omission and the change in its parameter values, but adds an $\mathsf{O}_p(1)$ variance, reducible by including $\{z_t\}$ in the model, offset by the parameter change and an estimation variance of $\mathsf{O}_p(T^{-1})$;
(*ivb*): slope estimation has $\mathsf{E}[(\widehat{\rho}-\rho_e)(x_T-\theta)]\simeq 0$ as $\mathsf{E}[\widehat{\rho}-\rho_e]=0$ by definition, and also $\mathsf{E}[x_T-\theta]=0$, but adds an estimation variance of $\mathsf{O}_p(T^{-1})$;
(*iva*): equilibrium-mean estimation also has $\mathsf{E}[(1-\rho)(\widehat{\theta}-\theta_e)]=0$, with an estimation variance of $\mathsf{O}_p(T^{-1})$;
(*iii*): forecast-origin uncertainty, $\mathsf{E}[\rho(\widehat{x}_T-x_T)]$, is zero only if the forecast origin is unbiasedly estimated, and will have a variance $\mathsf{O}_p(1)$ unless known for certain;
(*iib*) slope mis-specification again has $\mathsf{E}[(\rho-\rho_e)(x_T-\theta)]=0$, and adds an $\mathsf{O}_p(1)$ variance unconditionally;
(*iia*) equilibrium-mean mis-specification, where $\theta\neq\theta_e$, is possible if some in-sample location shifts were not modeled, but can be eliminated by a congruent model-selection procedure that removes breaks;
(*ib*) the slope change, $\mathsf{E}[(\rho^*-\rho)(x_T-\theta)]=0$ as $\mathsf{E}[x_T-\theta]=0$ irrespective of $\rho^*\neq\rho$;
(*ia*) that leaves just the equilibrium-mean change–which is the fundamental problem: $\theta^*\neq\theta$ induces systematic forecast failure.

Once in-sample breaks have been removed, even from good forecast origin estimates:

$$\mathsf{E}[\widehat{\epsilon}_{T+1\mid T}]\simeq(1-\rho^*)(\theta^*-\theta) \tag{36}$$

and that bias persists at $\widehat{\epsilon}_{T+2\mid T+1}$ etc., so long as (32) is used, even though no further breaks ensue. Keeping μ constant, but shifting ρ to ρ^*, induces a shift in θ to θ^*. The power of that insight is exemplified by:
(a) change *both* μ and ρ by large magnitudes but with $\theta=\theta^*$ generates an outcome that is isomorphic to $\mu=\mu^*=0$, so no break is detected as shown; and
(b) when $\mu=\mu^*=0$, then $\kappa\neq\kappa^*$ induces forecast failure by shifting θ even when z_{t-1} is correctly included and is known over the forecast horizon;
(c) as figure 6 shows, it is feasible to essentially replicate the same break by changing $\mu,\ \gamma$ and ρ in

many combinations, so no economists, economic agents, could tell what had shifted the outcome till long afterwards.

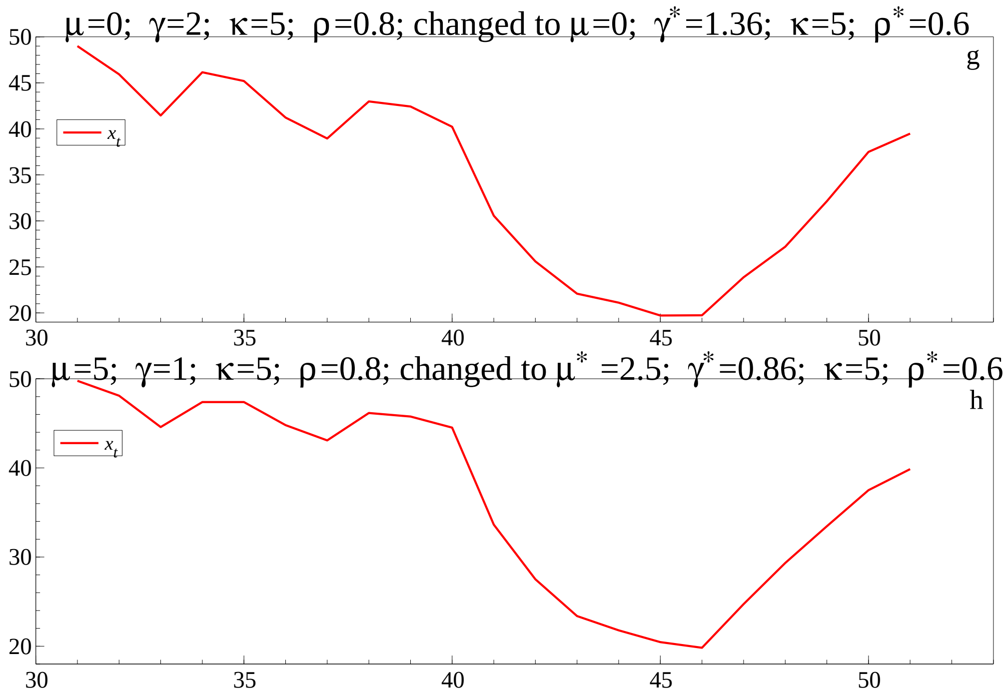


Figure 6: Many parameters shift

Such results apply to all equilibrium-correction models–they fail systematically when $\mathsf{E}[x]$ changes as the model's forecasts converge to θ irrespective of value of θ^*. However, when ρ^* is changed back to $\rho = 0.8$, the parameter values revert to those of the original DGP, so the old equilibrium is restored, and forecasts rapidly converge back to $\mathsf{E}[x]$. This suggests the original model also 'recovers' when the DGP reverts.

4.3 Robust forecasting devices

Robust forecasting devices may forecast better than any structural model in such shifting processes, as measured by root mean-square forecast errors (RMSFEs). Clements and Hendry (1998, 1999) provide a framework for analyzing the properties of forecasting models in wide-sense non-stationary processes, when the device being used does not coincide with the generating mechanism of the process (Clements and Hendry, 2008, provide a non-technical explanation). To summarize, they explain:
(a) the recurrent episodes of systematic mis-forecasting that have occurred historically;
(b) show that *ex ante* well-specified models need not forecast better than badly specified; since
(c) causally-relevant variables need not improve forecasting over irrelevant variables;
(d) the benefits of many of the empirical practices of forecasters, such as intercept corrections;
(e) show why pooling across a range of methods and models can be beneficial (see e.g., Hendry and Clements, 2004), but need not be unless carefully undertaken (see Hendry and Reade, 2008b);
(f) why forecasting devices that are robust to location shifts do not experience systematic forecast failure; and so

(g) dominate in forecasting competitions (such as Makridakis and Hibon, 2000: see Fildes and Ord, 2002, and Clements and Hendry, 2001a); hence
(h) why so-called 'naive devices' (simple adaptive devices like damped trend, differenced models, and exponentially weighted moving averages–EWMAs) can outperform; whereas
(i) equilibrium-correction models can experience systematic forecast failure (see Hendry, 2006).
The simplest robust forecasting device in Hendry (2006) is discussed in Hendry and Mizon (2012b) and Castle, Fawcett and Hendry (2011). Difference the mis-specified estimated model (32), so that $\Delta\widetilde{x}_{T+h|T+h-1} = \widehat{\rho}\Delta x_{T+h-1}$ or:

$$\widetilde{x}_{T+h|T+h-1} = x_{T+h-1} + \widehat{\rho}\Delta x_{T+h-1} \tag{37}$$

Despite using the 'wrong' $\widehat{\rho}$ for the first 5 forecasts, being incorrectly differenced, and omitting the relevant variable, nevertheless Figure 7 demonstrates that the robust forecasting device (37) avoids most of the last nine forecast errors for the DGP compared to (say) Panel c. The overall RMSFE is 6.6 for the in-sample model versus 5.5 for the robust forecasts here; but 3.8 versus 2.0 over the last nine forecasts, so is nearly halved, avoiding systematic failure.

Despite its non-specific assumptions, therefore, a theory of forecasting which allows for unanticipated structural breaks in an evolving economic mechanism for which the forecasting model is mis-specified in unknown ways can provide a useful basis for interpreting, and potentially circumventing, systematic forecast failure in economics. These implications seem to carry over to nowcasting.

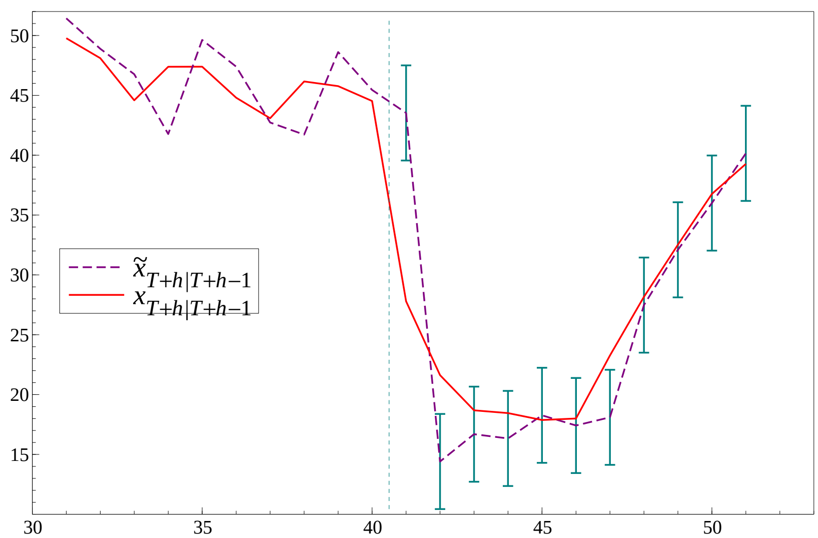


Figure 7: Differenced wrong model with changed ρ

A taxonomy of sources of forecast error for (37) clarifies why. Let $\widetilde{\epsilon}_{T+h|T+h-1} = x_{T+h} - \widetilde{x}_{T+h|T+h-1}$. We will use accurate $\widehat{x}_T$, with $\widehat{\rho} = \rho$ for simplicity but neither is crucial. Then from (37):

$$\widetilde{\epsilon}_{T+1|T} = (1-\rho^*)(\theta^* - \theta) - (1-\rho^*)(x_T - \theta) + \gamma^*(z_{T+1} - \kappa^*) - \rho\Delta x_T + \epsilon_{T+1}$$

so taking expectations, using $\mathsf{E}[x_T] = \theta$, for $h \geq 1$:

$$\mathsf{E}[x_{T+h}] \simeq \theta^* + (\rho^*)^h (\theta - \theta^*) \tag{38}$$

so:

$$\mathsf{E}[\Delta x_{T+h}] \simeq (1-\rho^*)(\rho^*)^{h-1}(\theta^* - \theta) \tag{39}$$

hence:

$$\mathsf{E}[\widetilde{\epsilon}_{T+1|T}] \simeq (1-\rho^*)(\theta^* - \theta) \tag{40}$$

which is equal to the in-sample DGP forecast bias.

However, at $T+2$ the differenced-device taxonomy becomes:

$$\widetilde{\epsilon}_{T+2|T+1} = -(1-\rho^*)(x_{T+1} - \theta^*) + \gamma^*(z_{T+2} - \kappa^*) - \rho\Delta x_{T+1} + \epsilon_{T+2} \tag{41}$$

so from (38) and (39):

$$\mathsf{E}[\widetilde{\epsilon}_{T+2|T+1}] \simeq (1-\rho^*)(\rho^* - \rho)(\theta^* - \theta) \tag{42}$$

Unless ρ^* has the opposite sign to ρ, there is a valuable offset from the $-\rho\Delta x_{T+1}$ component, helping explain the earlier forecast outcomes even though $\rho^* \neq \rho$.

Finally, at $T+3$:

$$\mathsf{E}[\widetilde{\epsilon}_{T+3|T+2}] = \rho^*(\rho^* - \rho)(1-\rho^*)(\theta^* - \theta)$$

which is close to zero, with $\rho^*(1-\rho^*) < 0.25$. Consequently, once $h-1 > 2$-periods after the break, using:

$$\widetilde{x}_{T+h|T+h-1} = x_{T+h-1} + \widehat{\rho}\Delta x_{T+h-1} \tag{43}$$

then $\widetilde{\epsilon}_{T+h|T+h-1} = (1-\widehat{\rho})\Delta x_{T+h-1}$ so:

$$\widetilde{\epsilon}_{T+h|T+h-1} = (1-\widehat{\rho})(\rho^* - 1)(x_{T+h-2} - \theta^* - \lambda^*(z_{T+h-1} - \kappa^*)) + (1-\widehat{\rho})\epsilon_{T+h-1} \tag{44}$$

as from equation (29):

$$\Delta x_{T+h-1} = (\rho^* - 1)(x_{T+h-2} - \theta^* - \lambda^*(z_{T+h-1} - \kappa^*)) + \epsilon_{T+h-1} \tag{45}$$

which therefore:
a] correctly reflects the new equilibrium through $(x_{T+h-2} - \theta^* - \lambda^*(z_{T+h-1} - \kappa^*))$;
b] includes the effect from z_{T+h-1} even though that variable was omitted from the forecasting device;
c] adjusts at the speed $(\rho^* - 1)$;
d] uses the well-determined in-sample estimate $\widehat{\rho}$, albeit that has shifted;
e] where $(1-\widehat{\rho})$ in (44) acts like a 'damped trend'.
Consequently, $\widetilde{x}_{T+h|T+h-1}$ uses: stochastic and deterministic post-break parameters; captures all unknown omitted variables; with consistent parameters and no estimation uncertainty; and no past data measurement errors. The main drawbacks are using an additional lag; being affected by the forecast-origin error $v_{T+h|T+h-1}$ (as does (33) of course); and retaining the past innovation ϵ_{T+h-1}. By way of contrast, $\widehat{x}_{T+h|T+h-1}$ based on the in-sample structural model suffers from breaks in both the stochastic and deterministic parameters; has an unknown number of omitted variables of unknown importance; with inconsistent parameter estimates and estimation uncertainty; contaminated by past data measurement errors. There are two major implications: first, there is no necessary connection between the in-sample verisimilitude of a model and its later forecasting performance; secondly, the conditional expectations at T (e.g., using the in-sample DGP) need not be the minimum RMSFE device for the outcome at $T+h$.

These taxonomy findings match previous graphs, and are little affected by estimating ρ. For the 'all parameters change' DGP, the robust device avoids almost all but the first forecast error in figure 8 panel

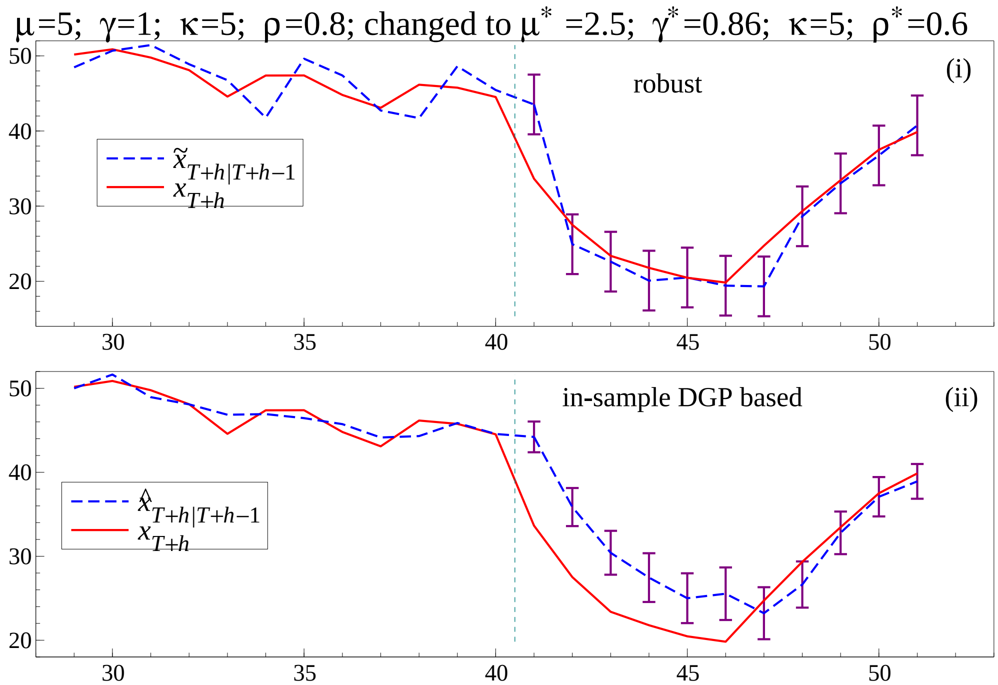


Figure 8: Differenced wrong model with changed ρ

(i), despite nearly all the parameters shifting, in stark contrast to the in-sample DGP forecasts in panel (ii), which show massive forecast failure for the first six forecast errors.

The robust device avoids systematic forecast failure after a location shift, at an insurance cost when no shifts occur. Returning to the first correct specification and compared to Figure 2 from (22), the RMSFE of (37) applied to that setting is 1.0 against the estimated DGP forecasts of 0.87.

4.4 Extending robust forecasting devices to I(1)

If the DGP produced non-trending I(1) data with an I(0) cointegrating relation given by $x_{t-1} - \theta - \lambda\,(z_t - \kappa)$, then:

$$\Delta x_t = (\rho - 1)\,(x_{t-1} - \theta - \lambda\,(z_t - \kappa)) + \epsilon_t \tag{46}$$

so merely rewriting the ADL would cover this additional non-stationarity. A drift term for a trend would add an intercept to (46), but otherwise, $\mathsf{E}[\Delta x_t] = 0$. Moreover, after a location shift, the robust forecast $\Delta\widetilde{x}_{T+h|T+h-1} = \Delta x_{T+h-1}$ is:

$$\Delta\widetilde{x}_{T+h|T+h-1} = (\rho^* - 1)\,(x_{T+h-2} - \theta^* - \lambda^*\,(z_{T+h-1} - \kappa^*)) + \epsilon_{T+h-1}$$

Thus, for $h \geq 2$, $\mathsf{E}[\Delta\widetilde{x}_{T+h|T+h-1}] \simeq 0$ as well.

4.5 Three empirical applications of forecasting

We will first illustrate the previous analysis by ex post forecasting Japanese exports over 2008(7)–2011(6), when they fell by more that 70% year on year, then comsider UK GDP over two recessions. In

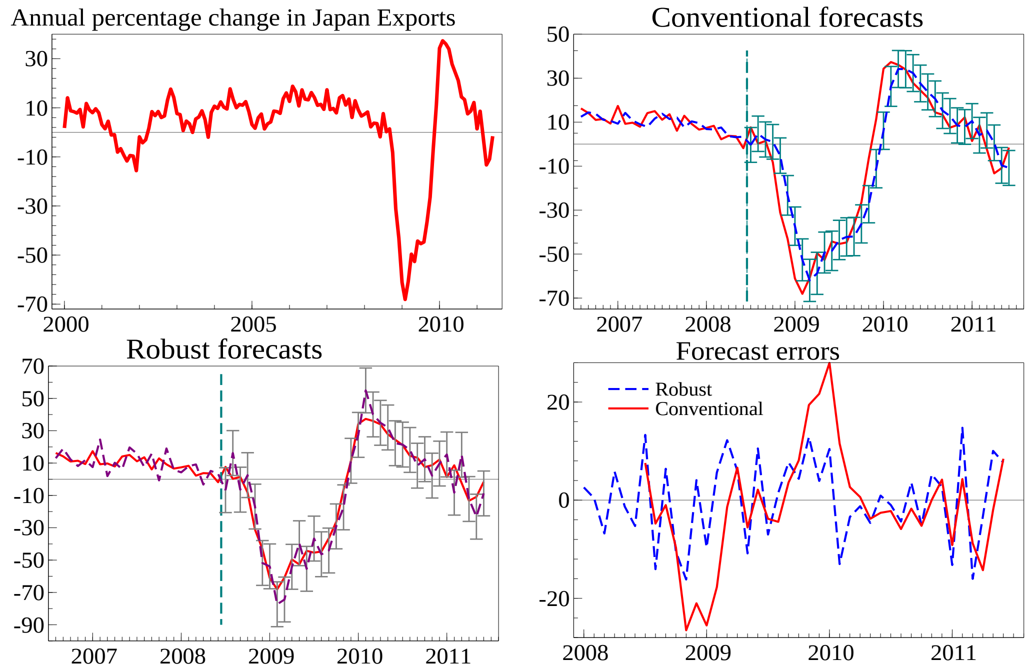


Figure 9: Forecasts of year-on-year changes in Japan's exports, 2008(7)–2011(6)

each case, we simply contrast an autoregressive model, often difficult to outperform in forecasting, with the robust variant of the same equation obtained by differencing.

For Japanese exports, the autoregressive model selected at 1% by *Autometrics* over 2000(1)–2008(6) was:

$$\widehat{y}_t = \underset{(0.091)}{0.68}\ y_{t-1} + \underset{(0.090)}{0.26}\ y_{t-3} + \underset{(0.04)}{0.12}\ 1_{2000(2)} + \underset{(0.04)}{0.12}\ 1_{2002(1)}$$

$$\widehat{\sigma} = 0.039 \quad \chi^2(2) = 0.65$$

$$\mathsf{F}_{\mathsf{het}}(4,95) = 1.02 \quad \mathsf{F}_{\mathsf{ar}}(6,92) = 0.97 \quad \mathsf{F}_{\mathsf{reset}}(2,96) = 2.1$$

where $1_{200z(x)}$ denote indicators for $200z(x)$. The model selected is almost 'robust' by having a near second unit root.

The corresponding robust device was the differenced device:

$$\widetilde{y}_t = y_{t-1} + 0.95\Delta y_{t-1} \qquad \widetilde{\sigma} = 0.0755$$

Their respective forecasts are shown in Figure 9.

This is a typical pattern: the robust device avoids forecast failure, at an insurance cost when no shifts occur. Overall, their comparative RMSFEs are $\mathbf{0.124}$ versus $\mathbf{0.098}$ as highlighted by Figure 10.

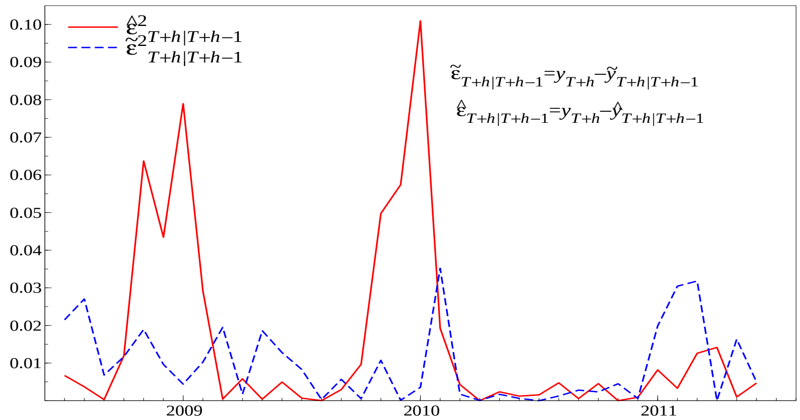


Figure 10: Squared forecast-error comparison

4.5.1 Forecasting UK GDP

The autoregressive model selected at 1% by *Autometrics* over 1989(2)–2007(4) was:

$$\begin{aligned}
\widehat{y}_t &= \underset{(0.083)}{0.537}\, y_{t-1} + \underset{(0.001)}{0.003} - \underset{(0.003)}{0.018}\, 1_{1990(3)} \\
\widehat{\sigma} &= 0.0032 \quad \chi^2(2) = 2.27 \quad \mathsf{F}_{\mathsf{ar}}(5,6) = 1.60 \\
\mathsf{R}^2 &= 0.51 \quad \mathsf{F}_{\mathsf{het}}(2,71) = 0.05 \quad \mathsf{F}_{\mathsf{reset}}(2,70) = 1.43
\end{aligned}$$

where $1_{1990(3)}$ is an indicator for $1990(3)$. The corresponding robust device was the difference:

$$\widetilde{y}_t = y_{t-1} + 0.5\Delta y_{t-1} \qquad \widetilde{\sigma} = 0.0053$$

Their respective forecasts across 2008–2011 are in Figure 11, and the RMSFEs are now 0.07 versus 0.04, so differencing nearly halves it, as panel d reveals.

4.5.2 Forecasts over an earlier debacle

HM Treasury also badly mis-forecast the UK recession in the early 1990s, so we reconsider that debacle in Figure 12. The forecasts from an autoregression are again outperformed by the robust differenced device. Note that an agency can wait until forecast failure has occurred before switching to a robust method, as the latter is no better at forecasting shifts: Castle, Fawcett and Hendry (2010), Castle *et al.* (2011) discuss the possibilities of forecasting location shifts.

5 Automatic model selection

In this section, we describe the basis of automatic model selection when there are more variables than observations.

Automatic model selection for nowcasting all disaggregates allowing for all the available information, multiple past breaks and contemporaneous location shifts can be undertaken following the approach described in Castle *et al.* (2012b), based on *Autometrics* with impulse saturation (see Doornik, 2009a, on

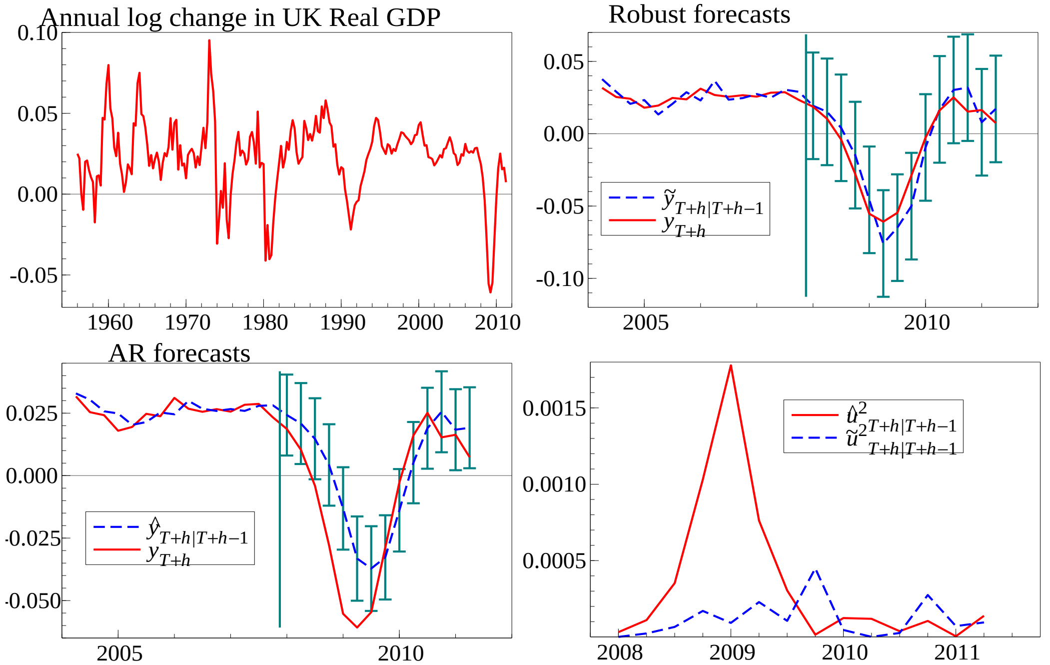


Figure 11: Data, forecasts and squared forecast errors over 2008–2011

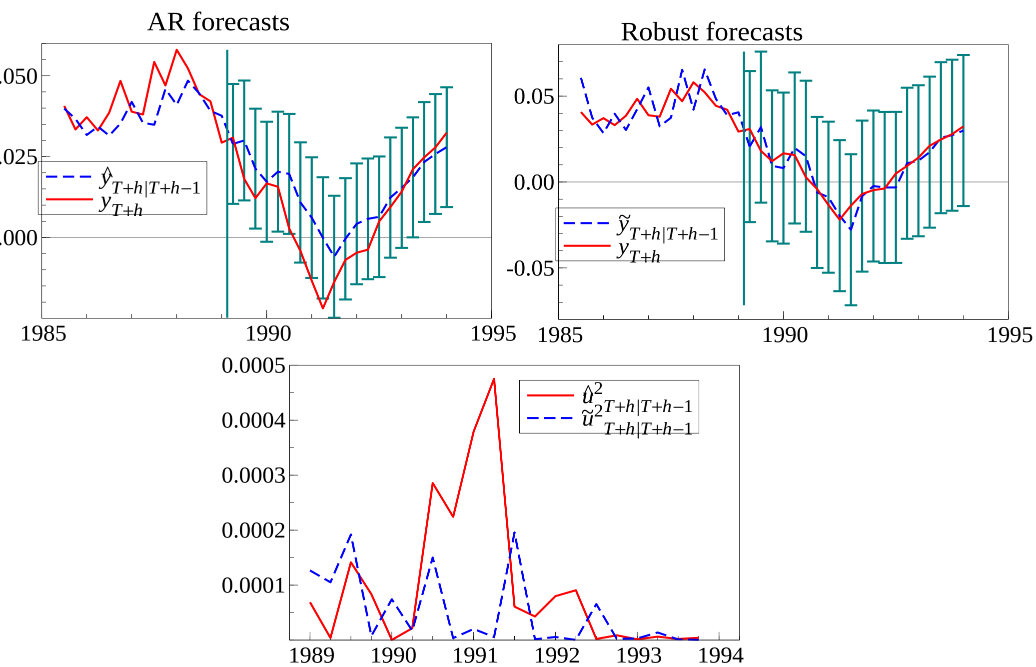


Figure 12: Forecast failure in the early 1990s

the former, and Hendry *et al.*, 2008, and Johansen and Nielsen, 2009, for the latter). The theory of reduction (see e.g. Hendry, 2009, for a recent exposition) explains the existence of a local data generation process (LDGP) for any subset of variables, and the relationship of any model to that LDGP. The idea behind general-to-specific (Gets) model selection is to locate a good approximation to that LDGP, characterized by its being a congruent representation and encompassing the evidence on the LDGP directly and via that embodied in other models. To do so, one commences from the most general model that nests all the candidate variables, their lags, functional form transformations and possible breaks, intrinsically leading to having more variables, N, than observations T (also see Castle *et al.*, 2012a, who consider using all the variables jointly with their principal components, with an extension by Castle *et al.*, 2012 to forecasting).

To model complex equations, we rely on general to specific modeling approaches (see Campos, Ericsson and Hendry, 2005). The unknown data generating process (DGP) is the underlying structure that creates the data. Empirical modeling will always deal with a subset of variables of the DGP, thus an important factor is the local data generating process (LDGP)—the generating process in the space of the variables under analysis: see Hendry (2009). The approach, therefore, is to construct a set of data based on broad theoretical assumptions, which nests the LDGP, then within this set, reduce the model from its general form down to a specific representation. This is a two step procedure. One: define a set of N variables that include the LDGP as a sub-model. Two: starting with that general model as a good approximation of the overall properties of the data, reduce its complexity by removing insignificant variables, while checking that at each reduction the validity of the model is preserved. This is the basic framework of *Gets* modeling.

This section introduces theoretical concepts of model selection, their use in mis-specification testing, followed by the introduction of impulse indicator saturation (IIS) and its generalized version. All these concepts are then united and applied through the automatic search algorithm *Autometrics*. The algorithm combines these features through automated selection based on *Gets* while handling more variables than observations with IIS for detecting breaks and outliers, and mis-specification testing.

The theory of reduction characterizes the operations implicitly applied to the DGP to obtain the local LDGP. Choosing to analyze a set of variables, denoted $\mathbf{y}_t, \mathbf{x}_{1,t}$, determines the properties of the LDGP, and hence of any models of $\mathbf{y}_t$ given $\mathbf{x}_{1,t}$ (with appropriate lags, non-linear transforms thereof, etc.). A congruent model is one that matches the empirical characteristics of the associated LDGP, evaluated by a range of mis-specification tests (see e.g., Hendry, 1995, and section 5.2). A model is undominated if it encompasses, but is not encompassed by, all other sub-models (see e.g., Mizon and Richard, 1986, Hendry and Richard, 1989, and Bontemps and Mizon, 2008).

5.1 Autometrics

Autometrics (see Doornik, 2009a) is the latest installment in the automated *Gets* methodology and is available in the *OxMetrics* software package. The algorithm is based on the following main components:[5]

1. GUM: The general unrestricted model (GUM) is the starting point of the search. The GUM should be specified based on broad theoretical considerations to nest the LGDP.

2. Pre-Search: prior to specific selection, a pre-search lag reduction is implemented to remove insignificant lags, speeding up selection procedures and reducing the fraction of irrelevant variables selected (denoted the gauge of the selection process). Pre-search is only applied if the number of variables does not exceed the number of observations ($N < T$)

[5]The following explanation is based on the exposition in Hendry and Pretis (2012).

3. Search Paths: *Autometrics* uses a tree search to explore paths. Starting from the GUM, *Autometrics* removes the least significant variable as determined by the lowest absolute t-ratio. Each removal constitutes one branch of the tree. For every reduction, there is a unique sub-tree which is then followed; each removal is back-tested against the initial GUM using an F-test. If back-testing fails, no sub-nodes of this branch are considered (though different variants of this removal exist). Branches are followed until no further variable can be removed at the pre-specified level of significance α. If no further variable can be removed, the model is considered to be terminal.

4. Diagnostic Testing: each terminal model is subjected to a range of diagnostic tests based on a separately chosen level of significance. These tests include tests for normality (based on skewness and kurtosis), heteroskedasticity (for constant variance using squares), the Chow test (for parameter constancy in different samples), and residual autocorrelation and autoregressive conditional heteroskedasticity. Parsimonious encompassing of the feasible general model by sub-models both ensures no significant loss of information during reductions, and maintains the null retention frequency of *Autometrics* close to α: see Doornik (2008). Both congruence and encompassing are checked by *Autometrics* when each terminal model is reached after path searches, and it backtracks to find a valid less reduced earlier model on that path if any test fails. This repeated re-use of the original mis-specification tests as diagnostic checks on the validity of reductions does not affect their distributions (see Hendry and Krolzig, 2003).

5. Tiebreaker: as a result of the tree search, multiple valid terminal models can be found. The union of these terminal models is referred to as the terminal GUM. As a tiebreaker to select a unique model, the likelihood-based Schwarz (1978) information criterion (SIC) is used, though other methods are also applicable, and terminal models should be considered individually.

In simulation experiments, models are primarily evaluated based on three concepts: gauge, potency and the magnitudes of the estimated parameters' RMSEs around the DGP values (see Doornik and Hendry, 2009). Gauge describes the retention of irrelevant variables when selecting (i.e., variables that are selected but do not feature in the DGP). Potency measures the average retention frequency of relevant variables (variables that are selected and feature in the DGP). Low gauge (close to zero) and high potency (close to 1) are preferred, as are small RMSEs.

The main calibration decision in the search algorithm is the choice of significance level α at which selection occurs. Selection continues until retained variables are significant at α, though it can be the case that variables in the final model are also retained at a level above α if removal leads to diagnostic tests failing. α is approximately equal to the gauge of selection. Further, the choice of diagnostic tests and lag length selection for residual autocorrelation and autoregressive conditional heteroskedasticity need to be set.

5.2 Mis-specification testing

Using a large number of variables with IIS (discussed in the next section) also provides a new view of model evaluation: to avoid mis-specification and non-constancy, start as general as possible within the theoretical framework, using all the available data unconstrained by $N > T$, (where N is the number of variables and T the number of observations) retaining the theory inspired variables and only selecting over the additional candidates.

Even so, this approach does not obviate the need to test the specification of the auxiliary hypotheses against the possibility that the errors are not independent, are heteroskedastic (non-constant variance) or non-normal, that the parameters are not constant, that there is unmodeled non-linearity, and that the conditioning variables are not independent of the errors. When $N << T$, the first five are easily tested in the initial general model; otherwise their validity can be checked only after a reduction to a feasible sub-model. Congruence is essential not only to ensure a well specified final selection, but also for correctly-calibrated decisions during selection based on Gaussian significance levels, which IIS will help ensure.

The model selection approach introduced here allows for more variables than observations to be used in modeling ($N > T$). For *Autometrics* this was first introduced through impulse-indicator saturation, which we now discuss, and has recently been extended to the general case.

5.3 Impulse-indicator saturation

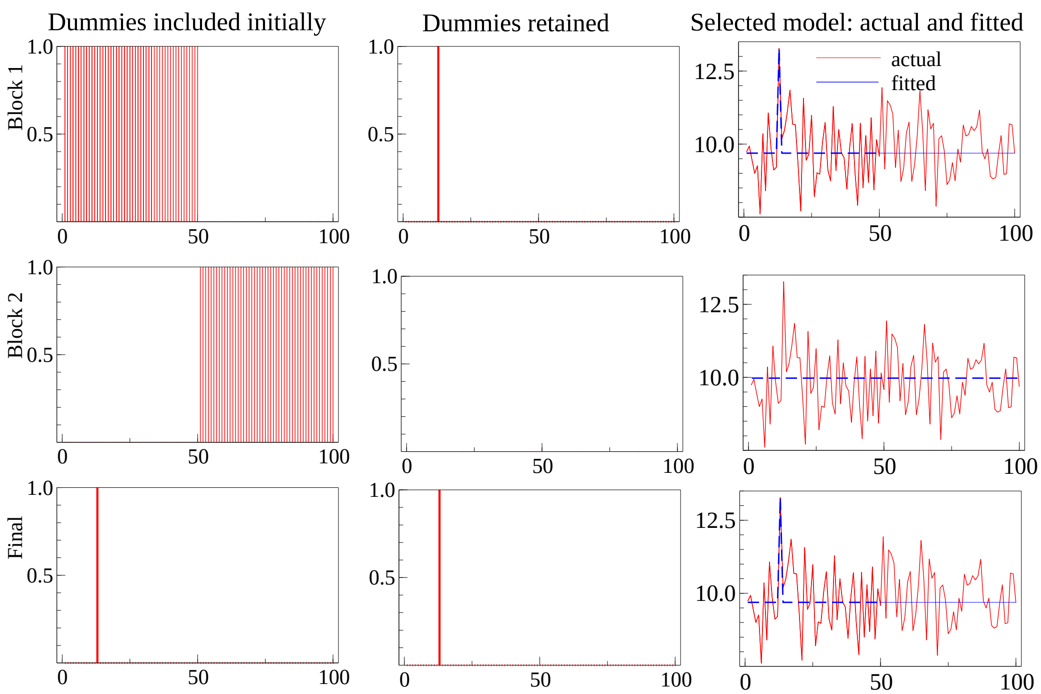


Figure 13: 'Split-sample' search under null

Autometrics handles the $N > T$ problem by a mixture of expanding and contracting searches that seek all the variables relevant at the chosen significance level α, set such that αN remains small (e.g., unity). Multiple breaks are tackled conjointly by impulse saturation (IS) which adds an impulse indicator for every observation: Hendry *et al.* (2008) establish one feasible algorithm, and derive the null distribution for an IID process, and Johansen and Nielsen (2009) generalize their findings to general dynamic regression models (possibly with unit roots), and show that there is a very small efficiency loss under the null of no breaks when αT is small, despite investigating the potential relevance of T additional variables. Castle *et al.* (2012b) examine the ability of IS to detect multiple breaks, and show it can find up to 20 breaks in 100 observations.

The numbers, timings and magnitudes of breaks in models are usually unknown, and are obviously unknown for unknowingly omitted variables, so a 'portmanteau' approach is required that can detect

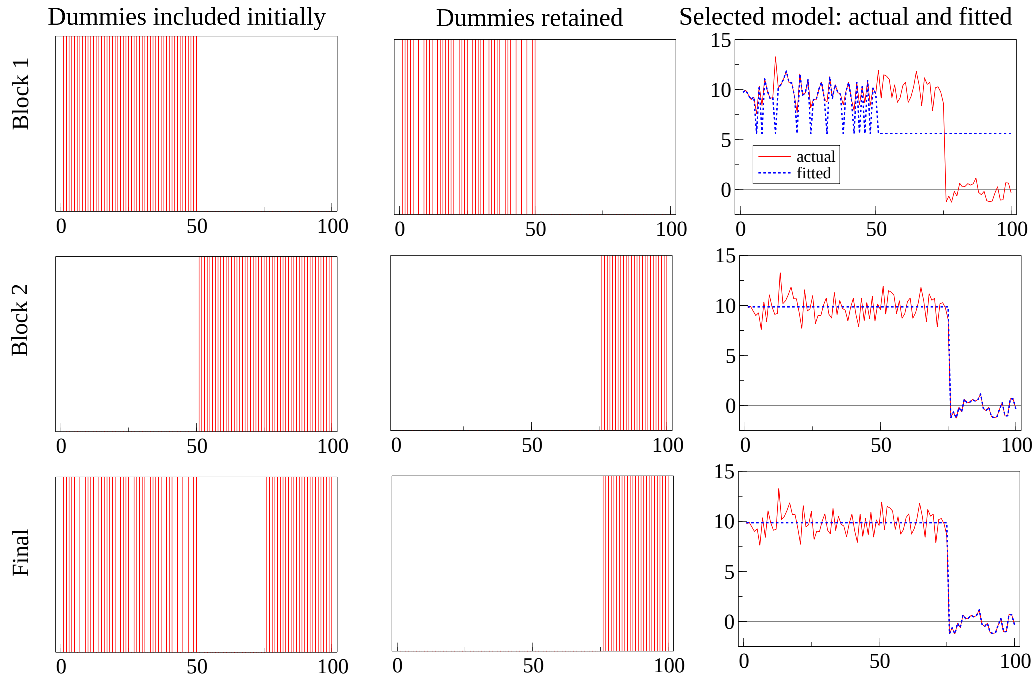


Figure 14: ‘Split-sample’ search under alternative

location shifts at any point in the sample while also selecting over many candidate variables. To check the null of no outliers or location shifts in a model, impulse-indicator saturation (IIS) creates a complete set of indicator variables $\left\{1_{\{j=t\}}\right\}$ equal to unity when $j = t$ and equal to zero otherwise for $j = 1, \ldots, T$, and includes these in the set of candidate regressors. Although this creates T variables when there are T observations, in the ‘split-half’ approach analyzed in Hendry *et al.* (2008), a regression first includes $T/2$ indicators. By dummying out the first half of the observations, estimates are based on the remaining data, so any observations in the first half that are discrepant will result in significant indicators. The location of the significant indicators is recorded, then the first $T/2$ indicators are replaced by the second $T/2$, and the procedure repeated. The two sets of significant indicators are finally added to the model for selection of the significant indicators. The distributional properties of IIS under the null are analyzed in Hendry *et al.* (2008), and extended by Johansen and Nielsen (2009) to both stationary and unit-root autoregressions.

Figure 13 illustrates the ‘split-half’ approach for $y_t \sim \mathsf{IN}\left[\mu, \sigma_y^2\right]$ for $T = 100$ selecting indicators at a 1% significance level (denoted α). The three rows correspond to the three stages: the first half of the indicators, the second half, then the selected indicators combined. The three columns respectively report the indicators entered, the indicators finally retained in that model, and the fitted and actual values of the selected model. Initially, although many indicators are added, only one is retained. When those indicators are dropped and the second half entered (row 2), none is retained. Now the combined retained dummies are entered (here just one), and selection again retains it. Since $\alpha T = 1$, that is the average false null retention rate.

We next illustrate IIS for a location shift of magnitude λ over the last k observations:

$$y_t = \mu + \lambda 1_{\{t \geq T-k+1\}} + \varepsilon_t \tag{47}$$

where $\varepsilon_t \sim \mathsf{IN}\left[0, \sigma_{\varepsilon}^2\right]$ and $\lambda \neq 0$. The optimal test in this setting would be a t-test for a break in (47) at $T-k+1$ onwards, but requires precise knowledge of the location-shift timing, as well as knowing that it is the only break and is the same magnitude break thereafter. Figure (14) records the behavior of IIS for a mean shift in (47) of $10\sigma_{\varepsilon}$ occurring at $0.75T = 75$. Initially, many indicators are retained (top row), as there is a considerable discrepancy between the first-half and second-half means. When those indicators are dropped and the second set entered, all those for the period after the location shift are now retained. Once the combined set is entered (despite the large number of dummies) selection again reverts to just those for the break period. Under the null of no outliers or breaks, any indicator that is significant on a sub-sample would remain so overall, but for many alternatives, sub-sample significance can be transient, due to an unmodeled feature that occurs elsewhere in the data set. Thus there is an important difference between 'outlier detection' procedures and IIS.

While IIS is perhaps surprising initially, many well-known statistical procedures are variants of IIS. The Chow (1960) test corresponds to sub-sample IIS over $T-k+1$ to T, but without selection, as Salkever (1976) showed, for testing parameter constancy using indicators. Recursive estimation is equivalent to using IIS over the future sample, and reducing the indicators one at a time. Johansen and Nielsen (2009) relate IIS to robust estimation, and show that under the null of no breaks, outliers or data contamination, the average cost of applying IIS is the loss of αT observations. Thus, at $\alpha = 0.01$, for $T = 100$ one observation is 'dummied out' by chance despite including 100 'irrelevant' impulse indicators in the search set and checking for location shifts and outliers at every data point. Retention of theory variables is feasible during selection with IIS, as is jointly selecting over the non-dummy variables, and IIS can be generalized to multiple splits of unequal size. While IIS entails more candidate variables than observations as $N + T > T$, selection is feasible as *Autometrics* undertakes expanding as well as contracting block searches (see next section). Non-linear model selection (including threshold models) is examined in Castle and Hendry (2011).

For a single location shift, Hendry and Santos (2010) show the detection power is determined by the magnitude of the shift τ; the length of the break interval $T - T^*$, which determines how many indicators need to be found; the error variance of the equation σ_{η}^2; and the significance level, α, where a normal-distribution critical value c_{α}, is used by the IIS selection algorithm. Castle *et al.* (2012b) establish the ability of IIS to detect multiple location shifts and outliers, including breaks close to the start and end of the sample, as well as correcting for non-normality. Figure 15 shows the application of the general autometrics algorithm to a trending process with four breaks of varying magnitudes over $1, \ldots, 10$; $40, \ldots, 45$; $75, \ldots, 90$; and $90, \ldots, 100$, to illustrate the ability of IIS to capture multiple breaks, at both the start and end of the sample.

5.4 General case of $N > T$

The idea of generalizing using more variables than observations from IIS to all forms of independent variables is introduced by Hendry and Krolzig (2005) as well as Hendry and Johansen (2012). Suppose there are $N = \sum_{j=1}^{N} n_j$ total regressors such that $N > T$ and $n_j < T$ for all j. Consequently the total number of variables N exceeds the number of observations T but total variables can be partitioned into blocks n_j each smaller than T. Their approach suggests randomly partitioning the set of variables into blocks of n_j, apply *Gets* to each block retaining the selected variables and crossing the groups to mix variables. The next step is to use the union of selected variables from each block to form a new initial model and repeat the process until the final union of selected variables is sufficiently small. *Autometrics* implements a variant of this algorithm to handle the general case of $N > T$.

In the general case of $N > T$ and IIS, *Autometrics* groups variables into two categories: selected and

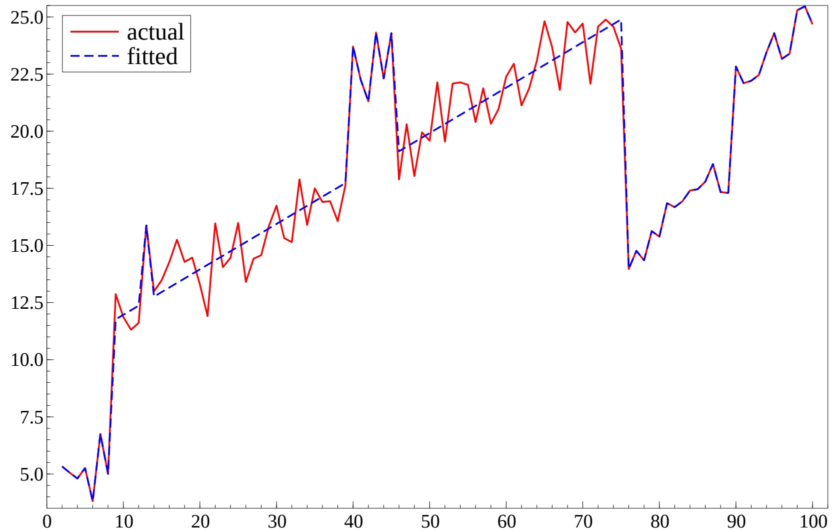


Figure 15: IIS outcomes for four location shifts

not selected (Doornik, 2007). Not currently selected variables are split into sub-blocks and the algorithm proceeds by alternating between two steps: first, the expansion step, selection is run over not-selected sub-blocks to detect omitted variables. Second, the reduction step, a new selected set is found by running selection on the system augmented with the omitted variables found in step one. This is repeated until the dimensions of the terminal model are small enough and the algorithm converges, so the final model is unchanged by further searches for omitted variables.

Autometrics has been applied successfully in a range of fields: see, for example, Hendry and Mizon (2011a) on US food expenditure, Bardsen, Hurn and McHugh (2010) on unemployment in Australia, and Castle *et al.* (2011a) for a comparison with other selection methods.

Nevertheless overall selections should be interpreted carefully. Successful identification of the underlying LGDP can be adversely affected by collinearity of the independent variables. Most simulations of *Autometrics* with large numbers of variables use orthogonal regressors, which makes selection easier. Furthermore, when $N > T$, in the block selection algorithm of *Autometrics*, adding or dropping a variable from the initial GUM may change the block partitioning of variables, so the selection is not invariant to the initial specification.

6 Producing 'good' nowcasts

In this section, we apply the preceeding analyses of forecasting and model selection to nowcasting an aggregate from disaggregates. The intent of Gets is to find 'good' LDGP models as defined in §5, even though Clements and Hendry (1998, 1999) show that one cannot prove that the pre-existing LDGP is the best model for forecasting.

A timely data source that provided an accurate 'measure' of a required aggregate variable is clearly preferable to seeking 'good forecasts' of it. Nevertheless, there will usually be a role for forecasting 'preliminary estimates', so good forecasting methods are likely to remain important, albeit that which methods might be 'best' is an empirical issue as seen in our extensive discussion of forecasting in §4.

Nowcasting is also partly a 'signal extraction' problem for missing data entering the aggregate, but if the first announcements can be systematically improved by forecasting them either directly or via the disaggregates, then the quality of the resulting aggregate is bound to be better. Three problems additional to the seven discussed in the Introduction inhibit achieving that outcome. All three considerations apply to nowcasting an aggregate either directly or via its disaggregates.

First, the objective functions of the users of a nowcaster's output are almost always unknown. A convenient approximation is to assume a quadratic loss, in which case the aim of nowcasting becomes to find a forecast $\widehat{y}_{T|T-\delta}$ of the aggregate y_T which solves:

$$\underset{\widehat{y}_{T|T-\delta}}{\operatorname{argmin}} \mathsf{E}_{\mathsf{D}_{y_T}} \left[y_T - \widehat{y}_{T|T-\delta} \right]^2 \tag{48}$$

where $\widehat{y}_{T|T-\delta} = g_{T-\delta}(\widehat{\mathcal{J}}_{T-\delta})$, and $g_{T-\delta}(\cdot)$ is the relevant function of the available measured information set, $\widehat{\mathcal{J}}_{T-\delta}$, that estimates the actual information, $\mathcal{I}_{T-\delta}$, on which y_T depends via $y_T = f_T(\mathcal{I}_{T-\delta})$. We have dated information as $T-\delta$ for $\delta > 0$, since some evidence must be unavailable at T to necessitate nowcasting. To proceed, we will assume that (48) does indeed provide the objective function, as it is difficult to see why either direction of asymmetry should dominate.

The second difficulty is that the information to be included in $\widehat{\mathcal{J}}_{T-\delta}$ is also unknown, and could comprise just the history of the relevant series (via a univariate time-series model), current and past data on other related time series, past information about revisions etc. We address this issue by proposing automatic model selection either directly for the aggregate itself, or for nowcasting the disaggregates allowing for all the available information, multiple past breaks and contemporaneous location shifts as in Castle *et al.* (2012b), then relating the aggregate to its own past and all the disaggregates as in Hendry and Hubrich (2011). Section 6.1 considers relating aggregates to disaggregates using automatic model selection (which section 5.1 discussed).

The third, and most serious, problem is obtaining $\widehat{y}_{T|T-\delta}$. The conditional expectation $\mathsf{E}_{\mathsf{D}_{y_T}}[y_T|\widehat{\mathcal{J}}_{T-\delta}]$ of y_T given $\widehat{\mathcal{J}}_{T-\delta}$ is only the solution to (48), namely the minimum mean-square error predictor, when no breaks occur. Unfortunately, $\mathsf{E}_{\mathsf{D}_{y_T}}[\cdot]$ is not known at T either, since the statistical process generating $\{y_t\}$ is never known, and in economics is always wide-sense non-stationary, namely its distribution changes over time from both stochastic trends and location shifts, reflected in our notation $\mathsf{D}_{y_T}(\cdot)$. Thus, the form of the optimal predictor $\widehat{y}_{T|T-\delta}$ is never known , and $\mathsf{E}_{\mathsf{D}_{y_{T-\delta}}}[y_T|\widehat{\mathcal{J}}_{T-\delta}]$, which might be available, need not be a good forecast device when location shifts occur, as it is calculated over the wrong distribution: see e.g., Castle, Doornik, Hendry and Nymoen (2013). Thus, the realized mean square forecast error (MSFE) given (48) is:

$$\mathsf{E}_{\mathsf{D}_{y_T}} \left[f_T\left(\mathcal{I}_{T-\delta}\right) - g_{T-\delta}\left(\widehat{\mathcal{J}}_{T-\delta}\right) \right]^2 \tag{49}$$

leading to the following abbreviated taxonomy of forecast errors $\widehat{u}_{T|T-\delta} = y_T - \widehat{y}_{T|T-\delta}$:

$$\begin{aligned}
\widehat{u}_{T|T-\delta} &= f_T\left(\mathcal{I}_{T-\delta}\right) - g_{T-\delta}\left(\widehat{\mathcal{J}}_{T-\delta}\right) \\
&= f_T\left(\mathcal{I}_{T-\delta}\right) - f_{T-\delta}\left(\mathcal{I}_{T-\delta}\right) \text{–location shift (§4.3)} \\
&\quad + f_{T-\delta}\left(\mathcal{I}_{T-\delta}\right) - g_{T-\delta}\left(\mathcal{I}_{T-\delta}\right) \text{–model mis-specification (§4, §5.2)} \\
&\quad + g_{T-\delta}\left(\mathcal{I}_{T-\delta}\right) - g_{T-\delta}\left(\mathcal{J}_{T-\delta}\right) \text{–reduced information (§6.1)} \\
&\quad + g_{T-\delta}\left(\mathcal{J}_{T-\delta}\right) - g_{T-\delta}\left(\widehat{\mathcal{J}}_{T-\delta}\right) \text{–measurement error (§6.8)}
\end{aligned} \tag{50}$$

Every term is potentially non-zero and doubtless ever present, but the expected values of the terms from model mis-specification and reduced information (which is not information about future location shifts) are zero. Thus, location shifts (the focus above) and measurement errors (the focus of nowcasting) seem likely to be the most pernicious problems when they occur.

6.1 Nowcasting aggregates from disaggregates

The variable of interest, y_T, say GDP, is an aggregate variable, comprising $y_t = \sum_{i=1}^{N_T} w_i y_{i,t}$ where $y_{i,t}$ are the disaggregates and w_i are the weights, which could be changing over time, as could the number of disaggregates. Data are released with varying time delays, such that at time T some components of y_T will be observed and some will be unavailable until $T+\delta$. For $y_{i,T}$, the $i = 1, \ldots, J_T$, components are known at T and the $i = J_T + 1, \ldots, N_T$, components are unknown at T. The two sets are not uniquely defined, so an individual element can switch back and forth between the sets over time depending on how the information is accrued, and J_T is not fixed. Hence, a forecasting strategy needs to be flexible enough to allow for changes in the timing of releases. As some components of y_T are unknown at T, a nowcast is computed. There are three alternative methods, see Hendry and Hubrich (2011), which we first describe, then address selection and estimation in §6.5:

I Forecast the aggregate using only aggregate information, some of which perforce must be lagged one period:

$$\widehat{y}_{T|T-\delta^*} = f\left(\mathbf{Y}^s_{T-1}, \mathbf{Z}^s_{T-\delta^*}\right) \tag{51}$$

where $\mathbf{Y}^s_{T-1} = (y_{T-1}, \ldots, y_s)$, and $\mathbf{Z}^s_{T-\delta^*} = (\mathbf{z}_{T-\delta^*}, \ldots, \mathbf{z}_s)$ is a vector of current and lagged conditioning variables such as surveys or leading indicators, which may be more recent than the latest aggregate observation, with $\delta^* \leq \delta \leq 1$.

II Forecast those disaggregates that are unknown at T:

$$\widetilde{y}_{i,T|T-\delta^*} = f\left(g\left(\mathbf{y}^0_{J_T,T}\right), \mathbf{Y}^s_{T-1}\mathbf{Z}^s_{T-\delta^*}\right), \quad i = J_T + 1, \ldots, N_T \tag{52}$$

where $g(\cdot)$ is a function, perhaps just a subset, of the known data in $\mathbf{y}^0_{J_T,T} = (y_{1,T} \ldots y_{J_T,T})$. Then the forecasts of the $\widetilde{y}_{i,T|T-\delta^*}$ could be aggregated with the known data:

$$\widetilde{y}_{T|T} = \sum_{i=1}^{J_T} w_i y_{i,T} + \sum_{i=J_T+1}^{N_T} w_i \widetilde{y}_{i,T|T-\delta^*} \tag{53}$$

Typically the weights are taken as given but could also be forecast; we abstract from the issue of weights in the subsequent analysis, as they are known with GDP data. Method [II] nests method [I], and could be applied even if $J_T = 0$, as would occur within a quarter, so high-frequency updates are possible in advance of measurements arriving. Since the measured variables in J_T change over time, §6.6 discusses how to specify $g(\cdot)$.

III Forecast the aggregate, conditioning on both aggregate and disaggregate information. There are two possible methods.

a. In order to apply standard forecasting procedures, a balanced panel is needed:

$$\overline{y}_{T|T-\delta} = f\left(\mathbf{Y}^s_{T-1}, \mathbf{y}^0_{J_T,T}, \mathbf{Z}^s_{T-\delta^*}\right), \quad \forall i \in N_T, \tag{54}$$

where the conditioning information includes information dated $T,\ T-\delta$ and previous, resulting in a δ-step ahead forecast.

b. Alternatively, an unbalanced panel can be used, conditioning on whatever information is also available at T:

$$\overline{\overline{y}}_{T|T} = f\left(\mathbf{Y}^{s}_{T-1}; \mathbf{y}^{0}_{J_T,T}, y_{J_T+1,T-\delta}, \ldots, y_{N_T,T-\delta}; \mathbf{Z}^{s}_{T-\delta^*}\right),$$

resulting in a large unbalanced panel in which there are missing observations at the end of the sample: see Wallis (1986) for a discussion of the 'ragged edge' problem.

Approach [I] has been the most common approach in the literature, with a focus on predicting the revision process at the aggregate level, see Lee, Olekalns and Shields (2008). However, the aggregate approach relies on a reduced information set, $\mathcal{J}_{T-1} \subset \mathcal{I}_{T-1}$, where $\mathcal{I}_{T-1}$ is the full information set including disaggregate information. As discussed in §3, unpredictability is relative to the information set used, so using a subset of information will result in less accurate predictions. However, although we do not use [I] here for the missing disaggregates, instead focusing on using the additional new disaggregate information, (51) enables *ex ante* forecasts of all disaggregates to be available before measured data arrive to be compared with those observed outcomes to test for any location shifts. When all disaggregates are observed, Espasa and Mayo (2012) discuss how to incorporate cointegrated and cyclical information about disaggregates to improve the aggregate forecasts relative to unrestricted estimation.

We now show analytically that when interest focuses on predicting the aggregate, then nothing is lost by doing so directly from disaggregate information, without predicting the disaggregates, i.e., method [III]. $N = 2$ suffices to illustrate the analysis, which generalizes to many components:

$$y_T = w_1 y_{1,T} + (1 - w_1)\, y_{2,T} \tag{55}$$

with weights w_1 and $w_2 = (1 - w_1)$, where for simplicity we take the weights to be constant over time, as in a simple sum aggregate. Assume that $J = 0$, so the contemporaneous disaggregates are not observed, and information is only available from $T - 1$. The DGP for the disaggregates is assumed to be:

$$y_{i,t} = \boldsymbol{\gamma}_i' \mathbf{x}_{t-1} + \eta_{i,t}$$

where $\mathbf{x}_{t-1} = (y_{1,t-1}, \ldots, y_{N,t-1}; \mathbf{z}_{t-1})'$ denotes all the available information, which includes the lagged disaggregates, so:

$$\mathsf{E}_{T-1}\left[y_{i,T} \mid \mathbf{x}_{T-1}\right] = \boldsymbol{\gamma}_i' \mathbf{x}_{T-1} \;\text{ for }\; i = 1, 2 \tag{56}$$

Aggregating the two terms in (56), delivers:

$$\mathsf{E}_{T-1}\left[y_T \mid \mathbf{x}_{T-1}\right] = \sum_{i=1}^{2} w_i \mathsf{E}_{T-1}\left[y_{i,T} \mid \mathbf{x}_{T-1}\right] = \left(\sum_{i=1}^{2} w_i \boldsymbol{\gamma}_i'\right) \mathbf{x}_{T-1} = \boldsymbol{\psi}' \mathbf{x}_{T-1} \tag{57}$$

say. Predicting y_T directly from $\mathbf{x}_{T-1}$ yields:

$$\mathsf{E}_{T-1}\left[y_T \mid \mathbf{x}_{T-1}\right] = \boldsymbol{\pi}' \mathbf{x}_{T-1} \tag{58}$$

Since the left-hand sides of (57) and (58) are equal:

$$\boldsymbol{\pi}' \mathbf{x}_{T-1} = \boldsymbol{\psi}' \mathbf{x}_{T-1} \tag{59}$$

so nothing is lost by predicting y_T directly, instead of aggregating component predictions once the same information set $\mathbf{x}_{T-1}$ is used for both, i.e. there is no benefit to method [IIIa] over [II].

Hendry and Hubrich (2011) show that this analysis generalizes to non-constant weights and changing parameters, and demonstrate that forecast-origin breaks equally affect combining disaggregate forecasts to forecast an aggregate as in (57) or forecasting the aggregate directly as in (58), so there is no benefit in MSFE terms from building models of disaggregates. However, that analysis assumes that all conditioning information is available at $T-1$, and does not make use of the contemporaneous information available at T that is an essential component of the nowcasting strategy. Discarding information available at T is particularly costly when breaks occur simultaneously across the disaggregates (or a subset) as this information can be used to ensure robust nowcasts despite the breaks. This implies that either [II] or [IIIb] is a preferable strategy, especially given a changing set in $\mathbf{y}^0_{J_T,T}$.

There is a range of methods for unbalanced panels in which all available data are used, including dynamic factor models, see Giannone *et al.* (2008) and Schumacher and Breitung (2008), and mixed frequency time-series models such as MIDAS, see Kuzin *et al.* (2009a). Marcellino and Schumacher (2010) combine factor models with MIDAS, and Ferrara *et al.* (2010) use non-parametric methods, based on nearest neighbors and on radial basis function approaches to nowcast unbalanced monthly data sets. However, these methods do not take account of structural breaks in the disaggregates, instead abstracting from wide-sense non-stationarity. Hence, the approach we propose is to use [II], but augmented by additional contemporaneous variables and location shift detection, the issues to which we now turn.

6.2 Common features

The disaggregate series are likely to have common components, such as cycles and trends, as well as idiosyncratic elements, so it can be useful to include these in the GUM: see Espasa and Mayo (2012). As N_T is large, it may be difficult to identify (say) cointegrating relationships across all disaggregates, but factors can proxy missing cointegration information. There are many ways to obtain the factors f_i for $i=1,\ldots,q$: Stock and Watson (2002b) suggest using static principal component analysis applied to $\mathbf{y}_{d,t}=(\mathbf{y}_{1_t},\ldots,\mathbf{y}_{N_t})'$:

$$\mathbf{f}_t=\widehat{\mathbf{G}}'\mathbf{y}_{d,t}$$

where the set of q factor loadings is collected in $\widehat{\mathbf{G}}$, the $N_T\times q$ matrix of eigenvectors corresponding to the q largest eigenvalues of the sample covariance matrix $\widehat{\boldsymbol{\Omega}}_{y_d}$. Alternatively, Forni *et al.* (2005) propose a weighted version of the static principal components estimator, where time series are weighted according to their signal-to-noise ratio. Although there is no asymptotic theory for static principal components (PCs) computed from I(1) data, viewed as data-based orthogonal transforms, PCs seem to capture the main aspects of common trends and cycles (a result first described for aggregates by Stone, 1947), so are probably sufficient for the purpose of nowcasting currently missing disaggregates. As the PCs mainly capture long-run relationships, the available information on the J_T known disaggregates at T will not contribute much, so the PCs can be calculated over the full sample to $T-1$.

In practice, we propose separating the disaggregates into blocks corresponding to groups of variables that exhibit common trends or cycles, as in section 7. The factors would be obtained for each block, f_i^j, where $i=1,\ldots,q$ denotes the factor and $j=1,\ldots,b$ denotes the block $\mathbf{y}_{d,j}$ where $\mathbf{y}_d$ is divided into b subsets, but all factors could be included in the GUM. A key innovation is that both individual explanatory variables and the factors are included jointly, in contrast to much of the dynamic factor models literature. This is feasible due to the ability of *Autometrics* to handle more variables than observations, and hence perfect collinearity, as described in §5. If factors are helpful in explaining movements in the disaggregates, they should be retained in the selected model.

6.2.1 Changing collinearity in nowcasting

Following the analyses in Clements and Hendry (1998, 2005), consider a linear, constant-parameter DGP where the first equation can be written for nowcasting as the contemporaneous model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t \tag{60}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$, assuming the simplest marginal process for illustration:

$$\mathbf{x}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \boldsymbol{\Sigma}\right] \tag{61}$$

independently of $\{\epsilon_t\}$, where $\boldsymbol{\Sigma} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'$ with $\mathbf{H}'\mathbf{H} = \mathbf{I}_n$ in-sample, and $\mathbf{z}_t = \mathbf{H}'\mathbf{x}_t$ so that:

$$\mathsf{E}\left[\mathbf{f}_t\mathbf{f}_t'\right] = \mathbf{H}'\mathsf{E}\left[\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{H} = \mathbf{H}'\boldsymbol{\Sigma}\mathbf{H} = \boldsymbol{\Lambda}$$

leading to the principal components (or factor) representation of (60):

$$y_t = \mathbf{x}_t'\mathbf{H}\mathbf{H}'\boldsymbol{\beta} + \epsilon_t = \mathbf{f}_t'\boldsymbol{\delta} + \epsilon_t \tag{62}$$

For simplicity we ignore sampling variability in estimating the principal components.

After time T, $\boldsymbol{\Sigma}$ changes to $\boldsymbol{\Sigma}^*$:

$$\mathsf{E}\left[\mathbf{x}_{T+1}\mathbf{x}_{T+1}'\right] = \boldsymbol{\Sigma}^* = \mathbf{H}'\boldsymbol{\Lambda}^*\mathbf{H}$$

with a concomitant shift to:

$$\mathsf{E}\left[\mathbf{f}_{T+1}\mathbf{f}_{T+1}'\right] = \boldsymbol{\Lambda}^*.$$

The 1-step ahead MSFE, $\mathsf{E}[\widehat{\epsilon}^2_{T+1|T}]$, for known regressors, after estimating either (60) or (62) is:

$$\sigma_\epsilon^2\left(1 + T^{-1}\mathsf{E}\left[\mathbf{x}_{T+1}'\boldsymbol{\Sigma}^{-1}\mathbf{x}_{T+1}\right]\right) = \sigma_\epsilon^2\left(1 + T^{-1}\mathsf{E}\left[\mathbf{f}_{T+1}'\boldsymbol{\Lambda}^{-1}\mathbf{f}_{T+1}\right]\right) = \sigma_\epsilon^2\left(1 + T^{-1}\sum_{i=1}^{n}\frac{\lambda_i^*}{\lambda_i}\right) \tag{63}$$

In (61), let $\mathbf{x}_t' = (\mathbf{x}_{1,t}' : \mathbf{x}_{2,t}')$ of dimensions n_1 and $n_2 = n - n_1$ setting $\mathsf{E}\left[\mathbf{x}_{1,t}\mathbf{x}_{2,t}'\right] = \boldsymbol{\Sigma}_{12} = \mathbf{0}$ both to simplify the algebra and emphasize that the problem is not 'omitted variables bias'. Then:

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \text{ and } \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{22} \end{pmatrix} \tag{64}$$

so that

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{11}\boldsymbol{\Lambda}_{11}\mathbf{H}_{11}' & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{22}\boldsymbol{\Lambda}_{22}\mathbf{H}_{22}' \end{pmatrix}$$

which similarly must apply after the shift:

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11}^* & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^* \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{11}\boldsymbol{\Lambda}_{11}^*\mathbf{H}_{11}' & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{22}\boldsymbol{\Lambda}_{22}^*\mathbf{H}_{22}' \end{pmatrix}$$

Consider a forecasting model that only includes the subset $\mathbf{x}_{1,t}$, leading to the 1-step ahead forecast for known regressors:

$$\widetilde{y}_{T+1|T} = \mathbf{x}_{1,T+1}'\widetilde{\gamma}_1, \tag{65}$$

where:

$$\mathsf{E}\left[\widetilde{\boldsymbol{\gamma}}_1\right] = \boldsymbol{\beta}_1 = \boldsymbol{\gamma}_1 \tag{66}$$

and:

$$\mathsf{V}\left[\widetilde{\boldsymbol{\gamma}}_1\right] \approx T^{-1}\left(\sigma_\epsilon^2 + \boldsymbol{\beta}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2\right)\boldsymbol{\Sigma}_{11}^{-1}.$$

The forecast error $\widetilde{\epsilon}_{T+1|T} = y_{T+1} - \widetilde{y}_{T+1|T}$ is:

$$\widetilde{\epsilon}_{T+1|T} = \mathbf{x}_{1,T+1}'\left(\boldsymbol{\gamma}_1 - \widetilde{\boldsymbol{\gamma}}_1\right) + \mathbf{x}_{2,T+1}'\boldsymbol{\beta}_2 + \epsilon_{T+1} \tag{67}$$

with unconditional expectation $\mathsf{E}\left[\widetilde{\epsilon}_{T+1|T}\right] = \mathbf{0}$ and approximate MSFE:

$$\begin{aligned}\mathsf{E}\left[\widetilde{\epsilon}_{T+1|T}^2\right] &\approx \sigma_\epsilon^2 + \mathsf{E}\left[\mathbf{x}_{1,T+1}'\mathsf{V}\left[\widetilde{\boldsymbol{\gamma}}_1\right]\mathbf{x}_{1,T+1}\right] + \boldsymbol{\beta}_2'\mathsf{E}\left[\mathbf{x}_{2,T+1}\mathbf{x}_{2,T+1}'\right]\boldsymbol{\beta}_2 \\ &\approx \mathsf{E}\left[\widehat{\epsilon}_{T+1|T}^2\right] + T^{-1}\sigma_\epsilon^2\sum_{j=n_1+1}^{n}\left(\tau_{\gamma_j}^2 - 1\right)\frac{\lambda_j^*}{\lambda_j}\end{aligned} \tag{68}$$

dropping the term of O(T^{-2}), and using (63) where $\mathbf{H}_{22}\boldsymbol{\beta}_2 = \boldsymbol{\gamma}_2$ with:

$$\tau_{\gamma_j}^2 = \frac{T\gamma_j^2\lambda_j}{\sigma_\epsilon^2},$$

which is the non-centrality of the t^2 test of the null that $\gamma_j = 0$ in (60) for $j = n_1+1,\ldots,n$.

Since only $\mathbf{x}_{1,t}$ is included in (65), when $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ the corresponding factors must be $\mathbf{f}_{1,t} = \mathbf{H}_{11}\mathbf{x}_{1,t}$ so the orthogonalized estimated forecasting model is:

$$\overline{y}_{T+1|T} = \mathbf{f}_{1,T+1}'\overline{\boldsymbol{\theta}}_1$$

where $\mathbf{f}_{2,T+1}'\boldsymbol{\theta}_2$ is omitted with:

$$\mathsf{V}\left[\overline{\boldsymbol{\theta}}_1\right] \approx T^{-1}\left(\sigma_\epsilon^2 + \boldsymbol{\theta}_2'\boldsymbol{\Lambda}_{22}\boldsymbol{\theta}_2\right)\boldsymbol{\Lambda}_{11}^{-1}$$

and a forecast error $\overline{\epsilon}_{T+1|T} = y_{T+1} - \overline{y}_{T+1|T}$:

$$\overline{\epsilon}_{T+1|T} = \mathbf{f}_{1,T+1}'\left(\boldsymbol{\delta}_1 - \overline{\boldsymbol{\theta}}_1\right) + \mathbf{f}_{2,T+1}'\boldsymbol{\theta}_2 + \epsilon_{T+1} \tag{69}$$

having an unconditional expectation $\mathsf{E}\left[\overline{\epsilon}_{T+1|T}\right] = \mathbf{0}$ and approximate MSFE:

$$\begin{aligned}\mathsf{E}\left[\overline{\epsilon}_{T+1|T}^2\right] &\approx \sigma_\epsilon^2 + \mathsf{E}\left[\mathbf{f}_{1,T+1}'\mathsf{V}\left[\overline{\boldsymbol{\theta}}_1\right]\mathbf{f}_{1,T+1}\right] + \boldsymbol{\theta}_2'\mathsf{E}\left[\mathbf{f}_{2,T+1}\mathbf{f}_{2,T+1}'\right]\boldsymbol{\theta}_2 \\ &\approx \mathsf{E}\left[\widehat{\epsilon}_{T+1|T}^2\right] + T^{-1}\sigma_\epsilon^2\sum_{j=n_1+1}^{n}\left(\tau_{\theta_j}^2 - 1\right)\frac{\lambda_j^*}{\lambda_j}\end{aligned} \tag{70}$$

when:

$$\tau_{\theta_j}^2 = \frac{T\theta_j^2\lambda_j}{\sigma_\epsilon^2}.$$

These non-centralities are orthogonal combinations of those in (68), and so could differ substantially. Nevertheless, the most important implications of changes in collinearity in forecasting are that an estimated model can have a smaller MSFE than the estimated DGP when omitted regressors (factors) have $\tau^2 \leq 1$ despite being mis-specified, and conversely all regressors with $\tau^2 > 1$ should be included in the forecasting equation. Changes in collinearity can markedly increase the MSFE when $\lambda_j^* >> \lambda_j$, irrespective of including or excluding regressors with $\tau^2 \leq 1$, so omitting relevant collinear regressors or orthognal factors need not improve a forecasting model. Finally, orthogonalizing transforms 'to reduce collinearity' are not useful when collinearity changes, since (68) and (70) both depend on the same eigenvalue ratios, albeit with different non-centrality deviations from unity.

6.3 Using all available information

The problem with forecasts from lagged variables only is that additional information may be available through both other contemporaneous data and disaggregates known at T, $y_{i,T}$ for $i = 1, \ldots, J_T$, allowing for more rapid identification of outliers or location shifts. Hence, we propose an augmented forecast denoted, $\widetilde{y}_{i,T|T-\delta}$.

There are many possible sources of contemporaneous data, including up-to-date values for some of the relevant disaggregates, rapidly and frequently observed outcomes for variables such as retail sales; correlated variables like road traffic and air passenger numbers or energy consumption; surveys of consumers and businesses about their plans and expectations; and more recent innovations including Google (see Choi and Varian, 2012) and prediction markets. Models that exploit related series, possibly in combination with univariate time-series models, can help improve the accuracy with which any missing data are estimated, by adding the proxy as an explanatory variable in a model for the variable to be forecast. The advantage of *Autometrics* is that all covariate information can be included in the general model at the outset, and the data characteristics will determine whether the explanatory variables are relevant historically in explaining the individual disaggregate series. The model selection for the individual forecast models could be undertaken for every new release of data, which is feasible for automatic model selection. Hence, if a break is detected the model will be updated as soon as information on the break is available. The robustness of the model specification and parameter constancy can be tested alongside the break detection tests, but the distinction between internal and external breaks discussed in Castle *et al.* (2010) points to different approaches in even those two cases.

6.3.1 Surveys

Survey information could be used directly to modify estimates of the forecast origin values, or as possible additional regressors, or as part of a signal extraction approach to estimating missing data on the disaggregates, or as one of the devices to be pooled. We doubt their likely efficacy as leading indicators following the critiques in Diebold and Rudebusch (1991) and Emerson and Hendry (1996), since their *ex post* performance is usually superior to the *ex ante*, reflected in regular revisions of the indicator components of indexes. However, we test whether surveys are able to provide timely detection of structural breaks in the empirical example.

6.3.2 Covariate information

Variables like retail sales are observed more frequently and rapidly than aggregates like GDP, and while they include some of the information needed for expenditure measures of GDP, they are also potentially correlated with other variables that are only available with greater latency. For example, retail sales are published approximately 32 days after the end of the reference period, whereas consumption data, for which retail sales is a reasonable proxy, is only released approximately 64 days after the end of the reference period along with the first full GDP release (excluding the flash estimate). Other proxies, such as new passenger car registrations, construction output and industrial new orders are also available more rapidly than components of the expenditure or income based measures of GDP.

6.3.3 Google query data

Choi and Varian (2012) show that *Google Trends* data can help improve their forecasts of the current level of activity for a number of different US economic time series, including automobile, home, and retail sales, as well as travel behaviour. They add the associated Google query variables to simple linear seasonal-AR models to measure the additional 'predictive' power, and although the resulting models are used for forecasting, their focus is on 'predicting the present'.

In an *Autometrics* approach, a small subset deemed to be potentially relevant additional data series could be added to the information set for the aggregate under analysis (Doornik, 2009b, illustrates this approach), or subsets could be used when nowcasting each missing disaggregates. Furthermore, the volumes of Google data available could enable measures of variance via the realized volatility of intra-day activity, and even complete distributions via a non-parametric approach. The optimal level of aggregation for such data is an empirical question, and mixed frequencies of data could draw on a MIDAS-type of approach (see, *inter alia*, Ghysels *et al.*, 2007, Ghysels *et al.*, 2007, and Clements and Galvão, 2008). The selected model specifications could be maintained and updated within quarters, or the models could be re-selected intermittently if breaks have occurred.

6.3.4 Prediction market data

Prediction markets for the relevant measures could be another valuable source of ancillary information: see Wolfers and Zitzewitz (2004), Gil and Levitt (2007) and Croxson and Reade (2008) *inter alia*. These information markets, like Iowa Electronic Markets for political outcomes and Betfair for sporting competitions, are claimed to have more accurate predictions than polls, surveys, or expert judgement, and to have predicted high profile events well, such as the probability of President Obama's first victory, as in figure 16. However, it is unclear how well such markets forecast relative to robust econometric equations, as location shifts occur intermittently, even if the reported probabilities adjust rapidly to new information (see e.g., Hendry and Reade, 2008a).

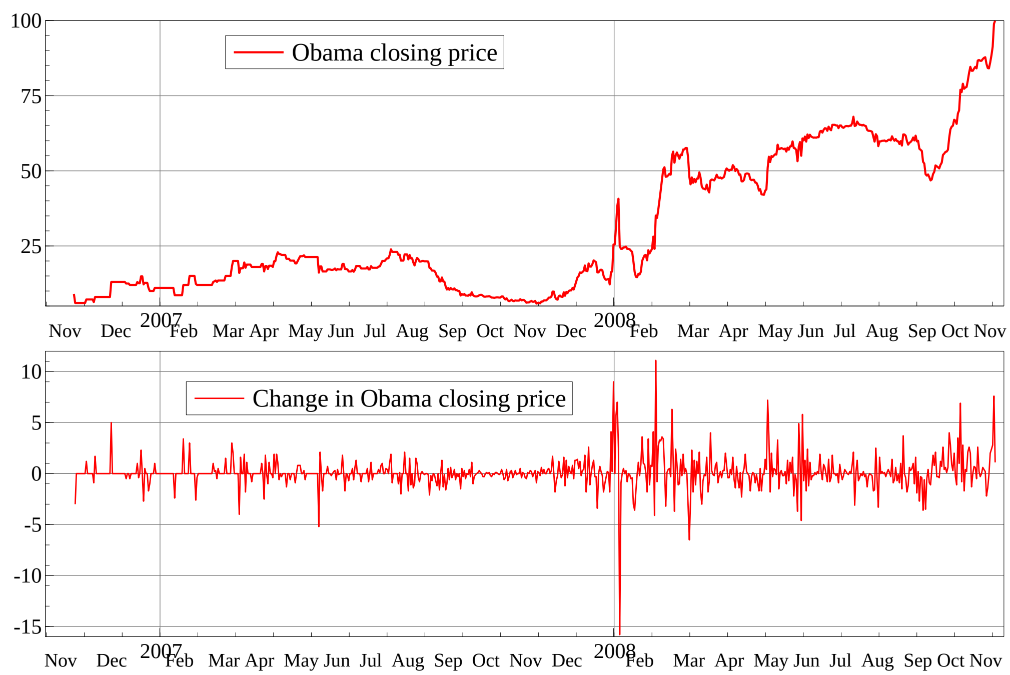


Figure 16: **Obama victory probabilities**

The graph shows that early forecasts are very poor, leading to forecast failure at long horizons. Certainly punters acquire information from many sources, including the economy, but also from poll outcomes and opinion surveys. There are large jumps at major events, such as winning the nomination etc., but adjustment is not monotonic even after as new shocks occur. Rapid updating seems to occur as news arrives, and accuracy improves at much shorter horizons, but even late in time, there are very inaccurate forecasts of the final probability. For example, the series predicts Obama's non-election as late as mid-September, some time after the structural break from Sarah Palin entering the race. Conversely, from mid-February onwards, most probabilities are above 0.5, so would be accurate for the outcome of an election victory if $\mathsf{p} > 0.5$ was the decision rule. Prediction markets inevitably have a termination point, by which date the probability has converged to either 0 or 1. Most graphs of the probabilities drift towards the actual outcome, but there can be 'last second' major breaks (e.g., an overtime goal in a soccer match), leading to dramatic switch in probabilities. Like most economic forecasts, future location shifts seem to be assumed absent.

Overall, however, as participants are risking their own money with their views, near their termination date, relevant prediction market real-time measures could be a useful addition to models measuring and nowcasting the present state of the economy. For example, the July 2009 Iowa Electronic Market was for bets on the Federal Reserve's Monetary Policy decisions, possibly revealing participants' implicit knowledge about the state of the US economy.

6.4 Seasonal adjustment

Seasonal adjustment (SA) of current (and recent past) data itself implicitly entails forecasting, as most SA procedures are long 2-sided moving-average filters, which have to be 'folded' at the end points of the sample, corresponding to extrapolating future data. Such 'forecasts' are often made by techniques that are different from the methods used for interpolating missing data in the disaggregated preliminary series. Conversely, if seasonally-adjusted disaggregated series underlie the published data, then the missing series must have been forecast in SA form. It is hard to ascertain whether such differences in methodology are of consequence relative to the other difficulties confronting the construction of preliminary data. Certainly, inconsistencies must arise when the forecasts implicit in the SA are not based on the same method as that used to fill in the missing data. Moreover, different outcomes will result if the data are first seasonally adjusted then forecast, or forecast then seasonally adjusted. Nevertheless, such errors are probably small relative to the mistakes in the initial data and any forecast numbers used to fill gaps.

6.5 Modeling the disaggregates

Models are required of all N_T disaggregates. As breaks may have occurred in the individual series over the sample period, it is important to account for location shifts in these models and we do so using IIS as described in §5.3. An illustrative in-sample general unrestricted model (GUM) for modeling the disaggregates is:

$$y_{i,t} = \boldsymbol{\gamma}'\mathbf{x}_{t-\delta} + \boldsymbol{\theta}'\mathbf{f}_{t-1} + \sum_{j=1}^{T-1} \varsigma_j 1_{t_j=t} + \rho_i y_{i,t-1} + \nu_t, \tag{71}$$

for $t = 1, \ldots, T-1$, and $i = 1, \ldots, N_T$. $\mathbf{x}_{t-\delta}$ denotes all the available information including other relevant lagged disaggregates and additional explanatory variables such as survey information, leading indicators etc. (as discussed in §6.3), $\mathbf{f}_{t-1}$ denotes a set of q latent common factors (discussed in §6.2), and $1_{t_j=t}$ is the set of $T-1$ individual impulse indicators taking the value unity at t and zero otherwise.

Selection is undertaken using *Autometrics* on each series' available sample (possibly $t = 1, \ldots, T-1$, but potentially shorter for recently introduced entities), and the resulting estimates are:

$$\widehat{y}_{i,t} = \widehat{\boldsymbol{\gamma}}'\mathbf{x}_t^* + \widehat{\boldsymbol{\theta}}'\mathbf{f}_{t-1}^* + \widehat{\rho}_i y_{i,t-1} + \widehat{\boldsymbol{\varsigma}}'\mathbf{d}_t \tag{72}$$

where $\mathbf{x}_t^*$ and $\mathbf{f}_{t-1}^*$ denote the retained variables and factors after selection and $\mathbf{d}_t$ collects the retained impulses from selection of (71). Loose significance levels could be used to select the forecasting model regressors in $\mathbf{x}_t^*$: Clements and Hendry (2005) show that all regressors with squared non-centralities of their t-statistics greater than unity should be retained. However, tight significance levels should be used to select the impulses as $\alpha(T-1)$ impulses will be retained on average under the null from (71). In practice, *Autometrics* only imposes one significance level, so selection could be undertaken using a 2-step procedure by appealing to the Frisch and Waugh (1933) theorem, first identifying the impulses and then selecting the regressors conditional on the impulses.

6.6 Nowcasting strategy

We can now integrate all the components of the Nowcasting strategy. Standing at time T, with all the available data for $T-1$ and any selected higher-frequency contemporaneous data, the models in (71) have been specified, selected and estimated as in (72). From these, forecasts of the outcome at T can be developed, either directly if there were no recent known location shifts, or robustly otherwise. We first take the former:

$$\widetilde{y}_{i,T|T-\delta} = \widehat{\boldsymbol{\gamma}}'\mathbf{x}_{T-\delta}^* + \widehat{\boldsymbol{\theta}}'\mathbf{f}_{T-\delta}^* + \widehat{\rho}_i y_{i,T-1} + \widehat{\boldsymbol{\varsigma}}' d_{T-1} \tag{73}$$

for $i = 1, \ldots, N_T$, where $\delta \simeq 0$. A 'flash forecast' of the aggregate y_T could be produced, and compared to other forecasts for early warning signs. That is why we suggest also calculating the robust forecast based on the first difference of (73):

$$\Delta\overline{y}_{i,T|T-\delta} = \widehat{\boldsymbol{\gamma}}'\Delta\mathbf{x}_{T-\delta}^* + \widehat{\boldsymbol{\theta}}'\Delta\mathbf{f}_{T-\delta}^* + \widehat{\rho}_i \Delta y_{i,T-1} + \widehat{\boldsymbol{\varsigma}}'\Delta d_{T-1} \tag{74}$$

where the parameter estimates are unchanged, and $\overline{y}_{i,T|T-\delta} = \Delta\overline{y}_{i,T|T-\delta} + y_{i,T-1}$. However, robust methods do not forecast location shifts, merely adapt rapidly after their occurrence, so have the same large forecast error when a break happens after a forecast is made. Thus, it is only necessary to switch after a break has been observed. A decision on that occurrence can be based on testing whether $\widehat{\varsigma}_{i,T-1}$ was significant, using a conservative significance level like 1.0%, noting that conventional tests have low or no power at sample end-points.

Once the first J_T components are observed, denoted $\widehat{y}_{i,T|T}$, their forecasts can be compared with these outcomes to evaluate forecast accuracy. Denote the forecast errors by:

$$\widehat{e}_{i,T|T} = \widehat{y}_{i,T|T} - \widetilde{y}_{i,T|T-\delta}, \qquad i = 1, \ldots, J_T,$$

Large values of $\widehat{e}_{i,T|T}$ relative to the standard error of the corresponding equation in (72) suggest recent location shifts, which then have implications for missing cognate, or related, variables. Major discrepancies between the forecast and the measured diasaggregates suggest a need to rapidly update for location shifts.

Because the J_T observed variables can change every period, they cannot be introduced as regressors at the model selection stage in (71), so we now discuss how $g(\cdot)$ in (52) is formed. Our approach is to use the information from the $\widehat{e}_{i,T|T}$ to modify the forecasts of the $N_T - J_T$ unknown disaggregates from

(73), albeit there may remain a role for judgment, depending on whether the forecast errors for the measured J_T variables are thought to be measurement mistakes (and hence transitory) or location shifts, and whether idiosyncratic or common and so correlated with the unknown disaggregates at T. Partition the J_T observed variables into related sets (e.g., automobile related, industrial production, prices, financial variables, interest rates, labour variables, housing market variables, leisure services, etc.), where there are common trends or cycles within subsets of disaggregates and hence close linkages, so breaks are likely to spread within blocks, and. Denote these groups by $J^*_{k,T}$ for $k = 1, \ldots, K_T$ say. Within each subset, calculate the average $\widehat{e}_{i,T|T}$ (possibly as a proportion of the level of the series at $T-1$) and denote this by $\overline{e}^*_{k,T|T}$. Then the proposed modification of the forecasting rule after substantial breaks for each partition is:

$$\widehat{y}_{h,T|T} = \widetilde{y}_{h,T|T-\delta} + \overline{e}^*_{h,T|T}, \quad h \in J^*_{k,T} \tag{75}$$

In essence, (75) is an intercept-corrected (IC) forecast analogous to setting the forecast 'back on track', but across disaggregate series as opposed to through time, which is the usual procedure. Consequently, in contrast to standard IC procedures that double the error variance, only a small increase should occur from (75) when each $J^*_{k,T}$ contains a number of series. This strategy is computationally feasible, allows for breaks in-sample and at the nowcast origin, while employing the full information set, but relies on the known disaggregates containing a correlated signal about the unknown disaggregates.

An alternative after the detection of a location shift is to robustify the forecasts of the missing series directly by using (74). However, while this will exploit the knowledge that a location shift has occurred, it will not use the information on its sign and magnitude provided by $\overline{e}^*_{k,T|T}$. That would not matter greatly for a shift that happened in an earlier period, but could be a substantive loss for a shift during $T-1$ to T. Such a robust device is akin to a 'Holt–Winters in-fill', but on the differences relative to the usual application, and while it does not impose the break magnitude as in (75), it doubles the innovation error variance. Both approaches could be used and pooled, there being considerable evidence on the benefits of pooling good methods (although not on 'pools' where some bad methods are included).

The final stage is to aggregate the nowcast series with data known at T as follows:

$$\widetilde{y}_{T|T} = \sum_{i=1}^{J_T} w_{i,T}\widehat{y}_{T|T} + \sum_{i=J_T+1}^{N_T} w_{i,T}\widetilde{y}_{i,T|T-\delta} \tag{76}$$

This approach of utilizing data on the subset of available disaggregates to construct the desired aggregate addresses the 'missing data', 'changing database', and 'break' problems, as impulse-indicator saturation removes past breaks, as well as including the additional contemporaneous variables. *Autometrics* writes the *Ox* code for each case analyzed, so it is easy to develop a general program for application to many variables.

6.7 Some simulation evidence

Simulation experiments of the alternative nowcasting approaches are reported in Kitov (2012). Here we briefly summarize the results in relation to the performance of the robust and intercept-corrected devices for permanent and transitory location shifts. The parameters in the simulations are summarized in table 1 for a VAR(1) DGP in 2 disaggregates. *Autometrics* selected the relevant subset of variables from the 100 initially included while also implementing IIS.

The results showed that intercept-corrected nowcasts can result in almost full adjustment for the bias created by a common location shift when contemporaneous information is used for correction. However, when the break is not transmitted forward, the intercept correction results in biased nowcasts that may

Table 1: Parameters in the simulations

$M = 1000$	number of simulated replications
$n = 100$	total number of exogenous candidate variables
$L = 10$	number of relevant exogenous explanatory variables
$T = 200$	pseudo in-sample sample size, forecast origin
$y_{i,0} = 0$	initial values for two disaggregates
$h = 4$	pseudo out-of-sample sample size
$5\sigma_{y_t}$	magnitude of the location shift
$\epsilon_{i,t} \sim \mathsf{IN}[0, 1]$	DGP errors, for $i = 1, 2$

differ by a large magnitude from those of the conditional model. Consequently, intercept correction transpires to be a high risk–high reward device, so should be used only when there is sufficient evidence of relatively permanent co-breaking in the series.

The robust nowcasts performed well on average, resulting in unbiased predictions under more scenarios than other methods. It resulted in a lower risk of getting the nowcasts badly wrong when the adjustment was unnecessary, yet the correction margins based on RMSE or MAE (mean absolute error) measures were much lower than those for the intercept-corrected nowcast in cases when both methods resulted in an improvement relative to the conditional model. Moreover, the robust device can successfully correct the nowcast in the worst-case scenario when breaks in series are of opposite directions.

6.8 Mis-measured disaggregates and location shifts

The most difficult setting is when the reported disaggregates are mis-measured at the same time as a location shift occurs. The former can happen because an unrepresentative sample of reports is returned to the statistical agency (e.g., large companies doing well when small are unknowingly doing badly or vice versa, etc.). If contemporaneously a location shift has occurred, inferring the missing disaggregates can be problematic. We address this problem by comparing the two forecasting devices with the measured disaggregate outcomes. A further useful comparator can be the change from the previous outcome for each disaggregate. There are five possible outcomes.

a] All three deliver the same general outcome. The most probably state is that the data disaggregates are accurate and there is no break, but an undetectable contemporaneous offsetting of mis-measured data by a similar magnitude shift in the opposite direction is conceivable. The missing data can best be in-filled using the average of the two forecasting devices.

b] The recorded data and the regression forecast are similar, but the robust device differs from both. This reveals that a location shift or a serious mis-measurement had happened in the previous period, but either the former was not carried forward (i.e., probably an earlier mis-measurement), or a shift has indeed happened but been mis-measured this period (noting that the regression forecast may not reflect a previous location shift). If investigation cannot clarify which is the correct state before the release date, the best that can be done is probably to in-fill using the regression forecast.

c] The recorded disaggregate data and the robust device are similar, but the regression forecast differs from both. This probably reveals that a location shift occurred in the previous period and the data reflect that but the contemporaneous variables did not correct the regression forecast enough. The missing data can best be in-filled using the robust forecasting device.

d] The robust device and the regression forecast are similar, but both differ from the recorded disaggregate data. If a location shift had occurred in the previous period, then it must have been captured by the contemporaneous variables in the regression forecast in order for it to match the robust device, but the most likely scenario is that either a new shift has happened that is not reflected in either forecast, or there is a serious mis-measurement. If the last can be excluded, the missing data can best be in-filled using the average of the two forecasting devices corrected by the average forecast error from the recorded disaggregates.

e] All three deliver different outcomes. Clearly the state of nature must have changed from the previous period, and had also changed then. Adjusting the robust device by the average forecast error from the recorded disaggregates may be the best that can be achieved if the situation cannot be clarified in time.

The next section presents an illustrative application of the nowcasting strategy.

7 Nowcasting UK GDP growth

This final section presents an empirical application of the nowcasting methodology to UK GDP growth over the last two decades. The apparent break in the GDP growth series, represented by the recession, is a good candidate for testing the nowcasting performance of *Autometrics* with IIS for break correction, given the recent poor performance of the GDP estimates produced by the ONS. In addition, it is interesting to check whether *Autometrics* can provide robust nowcasts for two periods with difference types of variability. To that end, the period spanning the recent turmoil is compared with a calmer period during the 1990s.

The conditioning set includes many monthly leading indicators, as well as preliminary GDP growth vintages. The GDP growth nowcasts are computed using bridge equations. To check for potential benefits of information about common breaks, leading indicators are grouped in blocks of correlated variables that are more likely to break simultaneously. *Autometrics* model selection with IIS then produces forecasts of the leading indicators for the missing observations at the end of the sample. A robust forecasting device for monthly variables is also tested. An analysis of common breaks within blocks of closely linked variables checks whether contemporaneous intercept correction could be applied to improve nowcasting performance. Finally, nowcasting models for GDP growth are selected using *Autometrics*.

Section 7.1 reviews the data, §7.2 describes the handling of the leading indicators, §7.3 outlines the nowcasting methodology, and §7.4 discusses the model specifications. Then sub-sections 7.5–7.7 discuss the estimates for vintage models, single-indicator models and augmented models respectively. Section 7.8 investigates contemporaneous breaks.

7.1 Data

The GDP growth series was downloaded from the ONS website,[6] and contains 76 quarterly observations for the period between 1993Q1 and 2009Q4. Seasonally-adjusted chained volume measures of GDP at constant 2008 prices were used. The latest GDP growth estimates, presumed to be the most accurate reported value to date, are the target variable, denoted by Δy_t, where y_t is the log of the GDP level and Δ is the first difference of log(GDP). The latest GDP growth vintage is an appropriate measure when the objective is to nowcast the actual change in output, as in Corradi, Fernandez and Swanson (2009) and Clements and Galvão (2009b).

Four additional estimates of GDP growth were obtained, all from the GDP Revisions Triangle Release (ABMI) reported by the ONS. These are three preliminary vintages: $\Delta y_{t_q}^{v_1}$, $\Delta y_{t_q}^{v_2}$, $\Delta y_{t_q}^{v_3}$ released

[6] http://www.ons.gov.uk/ons/datasets-and-tables/index.html

in the months directly following the reference quarter, and $\Delta y_{t_q}^{vf}$ corresponding to the 'final estimate' reported three years after the end of the reference quarter (further revisions in later periods produce the latest estimate Δy_{t_q}). Consequently, the last is the most accurate measure of GDP available to date. A summary of the GDP estimates, their release dates and information content is provided in table 2.

Table 2: GDP growth estimates, release dates and data completeness

Variable	Estimate	Release lag	Real data
$\Delta y_{t_q}^{v1}$	Preliminary, first estimate	3.5 weeks	44%
$\Delta y_{t_q}^{v2}$	Output, Income and Expenditure, second estimate	8 weeks	67%
$\Delta y_{t_q}^{v3}$	UK National Accounts, first final estimate	12 weeks	80%
$\Delta y_{t_q}^{vf}$	Final estimate 3 years later	3 years	
Δy_{t_q}	Latest available, most accurate estimate to date		

To see the differences between the preliminary and the final growth estimates, the least complete estimate, $\Delta y_{t_q}^{v1}$, and the latest available one, Δy_{t_q}, are plotted in figure 17. Although the patterns of growth are similar on average, the deviations between the two series are considerable, especially in periods of major changes. Until the start of the recession in 2008, GDP growth had been positive and relatively stable, with non-trending and mean-reverting properties. The two series nearly coincide after 2010 as preliminary measures have not yet been substantially revised, so the latest quarterly estimate deemed reliable enough here is 2009Q4.

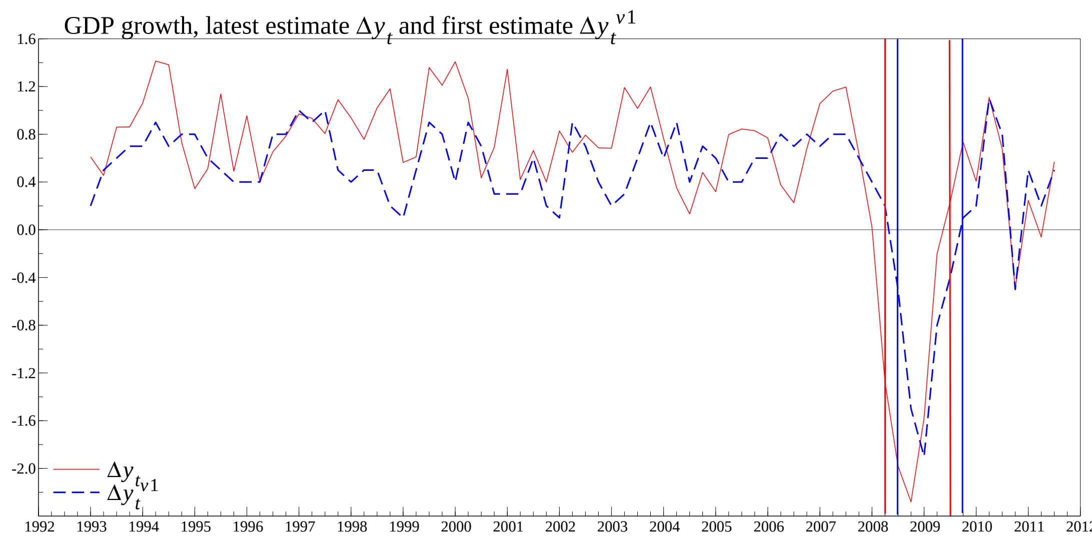


Figure 17: GDP growth in percentage terms based on the latest estimate Δy_t and the first preliminary estimate Δy_t^{v1} between 1993 and 2011

The first instance of negative growth was reported in 2008Q2, according to Δy_{t_q}. However, the recession was missed by preliminary estimates at all real-time vintages by one quarter. According to these measures, including $\Delta y_{t_q}^{v1}$, the recession did not start until 2008Q3, as can been seen on the plot. Thus, measurement errors during the recession were considerably larger than before. The recession started earlier than ONS reported for three consecutive National Accounts releases, and was only identified *ex* *post*. The measurement procedures failed to produce correct estimates of the output decline, with an estimation error of -1.5% at the first reported quarter of negative growth.

Further analysis of the revision process is presented in figure 18, which depicts the differences between Δy_{t_q} and the three preliminary vintages. The three revision processes are close, even for the final estimate $\Delta y_{t_q}^{v_f}$, noting that $\Delta y_{t_q}^{v_f}$ is only available until 2008Q3 and for 2008Q1–2008Q2 is equal to the latest estimate. All primary estimates tend to underestimate large positive levels of Δy_{t_q}, and to overestimate large negative ones, which happens in 70% of the cases. In general, even $\Delta y_{t_q}^{v_f}$ tends to underestimate the latest available estimate up to 2008Q3. The largest deviations from the most accurate estimates are reported during the recession, with two instances of 1.5% errors for the preliminary estimates.

Table 3: Revision to the latest and initial GDP growth estimate, mean error (ME) and standard deviation (SD) in percentage points for 1993Q1–2008Q3

	$\Delta y_t^{v_1} - \Delta y_t$	$\Delta y_t^{v_3} - \Delta y_t$	$\Delta y_t^{v_f} - \Delta y_t$
ME (%)	-0.152	-0.149	-0.113
SD (%)	0.474	0.435	0.309

	$\Delta y_t^{v_2} - \Delta y_t^{v_1}$	$\Delta y_t^{v_3} - \Delta y_t^{v_1}$	$\Delta y_t^{v_f} - \Delta y_t^{v_1}$
ME (%)	-0.011	0.003	0.040
SD (%)	0.083	0.116	0.379

Finally, table 3 reports average errors and standard deviations for the revisions compared to Δy_t, revealing systematic underestimation on average for all vintages, even including the most recent overestimation during the recession, which drives the overall mean error up from -0.200. As shown in figure 18, real-time vintages are not revised greatly compared to $\Delta y_t^{v_1}$. A noticeable, but unsystematic error, can be also seen for the final estimate, especially at the first instance of recession.

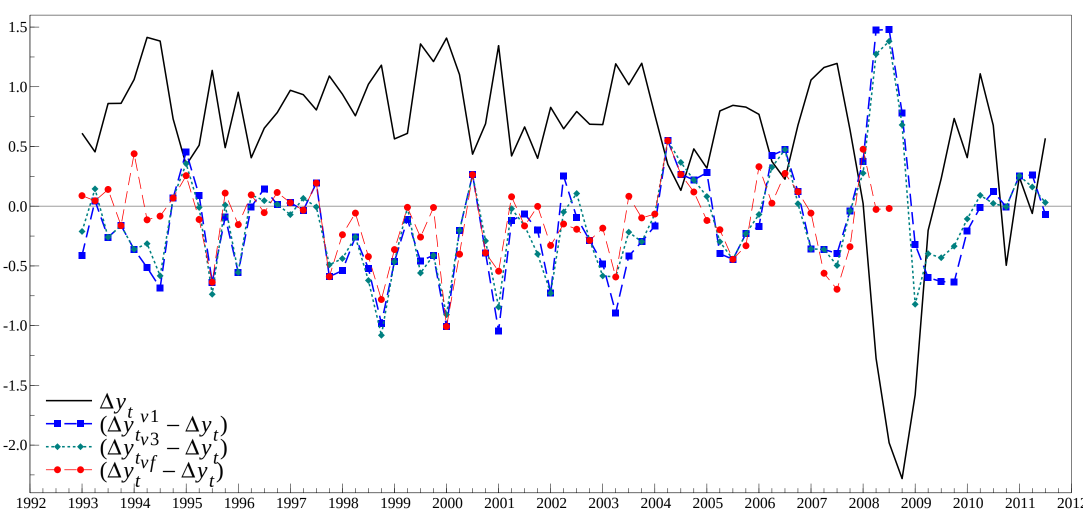


Figure 18: Revision process to the latest estimate of GDP growth, Δy_t, as compared with the $\Delta y_t^{v_1}$, $\Delta y_t^{v_3}$ and a revised GDP measure three years later, $\Delta y_t^{v_f}$

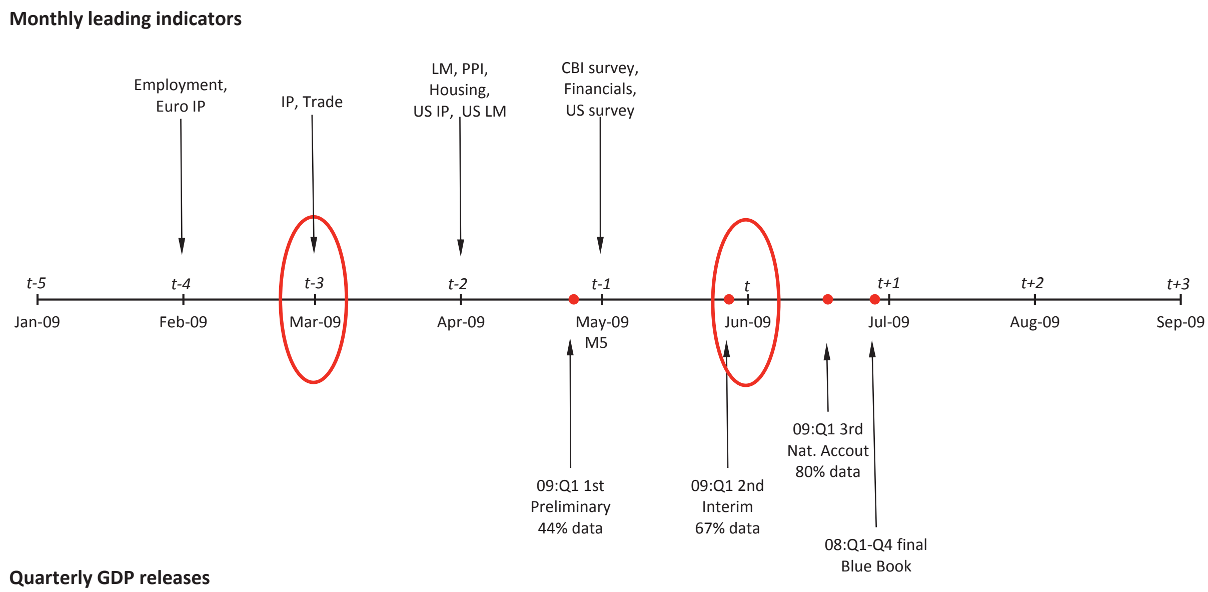


Figure 19: Pseudo release timing and structure of data vintages for the quarterly GDP measurements and monthly leading indicators

7.2 Leading indicators

The data set on the monthly disaggregated variables contains 60 series, over 1980Q1–2010Q2, and was kindly provided by Christian Schumacher: see Kuzin, Marcellino and Schumacher (2009b). All of these series had some merit in explaining GDP in previous studies, and include UK survey data, industrial production, labor market and employment statistics, trade variables, financial indicators and the corresponding series for the US and Europe. Leading indicators are released with different monthly lags relative to the current month t_m, where the actual days of publication differ throughout the month. In a real-time exercise, the information set can expand by a single variable and each new release can be treated as a nowcast origin, as in Bánbura *et al.* (2011). However, it suffices to fix the release lags relative to the current period for each variable, resulting in a constant structure for an unbalanced panel, where the pattern of missing observations at each point in time is the same. This produces a pseudo real-time data set, which in real time is only available at the end of each month.

The level series are non-stationary so data transformations are applied to produce variables that appear to be approximately I(0). For application of the common-break-correction devices, the leading indicators were grouped into 6 blocks, as in Castle *et al.* (2011), where the blocks should contain variables with varying release lag orders to aid timely detection of shifts. Suc variables should be highly correlated, increasing the likelihood of common breaks. Therefore, six blocks were formed by optimizing cross-correlations between the transformed series, so that within-block correlations are maximized, whereas between-blocks correlations are minimized, while trying to group variables with different release lags. The latter restriction is not always satisfied. For example, block 5 only contains financial variables that are available contemporaneously, but have negligible correlations with other variables and are therefore grouped separately.

The actual timings of data releases for the leading indicators and GDP estimates are presented in figure 19. Suppose that the current period $t = t_{m_6} = t_{q_3}$ is June. Then, as can be seen in the upper half of the figure, the latest UK and US soft data together with the financial variables are available for the previous month, i.e., with a one-month lag. The remaining leading indicators are released with longer lags. For instance, employment data and European industrial production are only available for February. Three real-time vintages of GDP growth are also shown in the figure, and for the illustrated example are all available for the first quarter.

7.3 Nowcasting methodology

Nowcasting GDP growth is performed using the bridge equations framework. At the first stage, direct forecasts of the missing observations for all leading indicators are computed within blocks of correlated variables. The forecasts are produced for the last 24 months of each corresponding sub-sample, that is for 1999M1–2000M12 and 2008M1–2009M12 respectivley. Individual models are selected recursively using *Autometrics* with IIS at a 1% significance level. For a monthly variable z_{i,t_m} in block $k = \{1, \dots, 6\}$, available with a lag l_i, such that the latest observation is $z_{i,t_m - l_i}$, the following GUM fully saturated by impulse indicators is formed from variables of the same block available in-sample:

$$z_{i,t_m} = \sum_{i=1}^{b_k} \sum_{j=1}^{12} \beta_{i,j} z_{i,t_m - l_i - j} + \sum_{t=1}^{t_m} \zeta_{i,t} 1_{i,t} \tag{77}$$

where b_k is the total number of variables in block k and 12 is the largest lag entering the GUM. All monthly indicators are forecast within blocks, that is for $1 \leq i \leq b_k$. One, two or three periods ahead forecasts are computed depending on the corresponding release lag $l_i \in \{1, 2, 3\}$ to correct for the ragged edges. The earliest lag allowed to enter the GUM varies among variables and corresponds to the release delay l_i, so that all information that would have been available in real-time at the forecast origin is included, so contemporaneous observations for variables available at t_m are included in a model for all other series observed with a lag, but not vice versa. Forecasting models are also estimated for the variables released with no delays, resulting in one step ahead forecasts. Break detection is performed simultaneously with model selection and significant contemporaneous impulses are retained. This is done for a later analysis of common breaks at the nowcast origin.

The second stage of the nowcasting procedure involves the formulation of the bridge equation linking the quarterly GDP growth with the monthly leading indicators. In order to check whether monthly inflow of information can monotonically improve nowcasting accuracy, three nowcast origins are compared. These are: $h^1_{t_{q_1}}$ corresponding to the second month of the quarter, $h^2_{t_{q_1}}$ corresponding to the third month of the quarter and $h^3_{t_{q_1}}$—one month after the reference quarter. The structure of nowcast vintages is shown in table 4, along with the releases of the real-time quarterly GDP growth estimates. Note that at $h^3_{t_{q_1}}$ the first preliminary estimate is available. The latter is intentionally included to test whether the flash estimates contain relevant information about actual growth in real time.

Provided that the period before the recession is characterized by mean reverting growth with low volatility, it is conjectured that the model selected for this sub-sample will differ substantially from the model selected for the recession. In order to check this, the whole available period is divided into two sub-samples. The first sub-sample contains 32 observations and spans the period between 1993Q1 and 2000Q4. The second sub-sample covers 36 observations over 2001Q1–2009Q4. For each period, the last eight quarters are nowcast as noted above.

The mixed-frequency data issue is tackled explicitly within the bridge equations framework by transforming the unmodeled monthly variables to produce a consistent quarterly panel so that a nowcasting regression for the quarterly data could be specified directly. Consequently, to use the full information

Table 4: Nowcast horizons and GDP releases in real time. Each entry corresponds to a nowcast/release of the quarterly GDP growth in a quarter t_{q_i} available in month t_{m_j}

	t_{m_1}	t_{m_2}	t_{m_3}	t_{m_4}	t_{m_5}	t_{m_6}	t_{m_7}	t_{m_8}	t_{m_9}	$t_{m_{10}}$	$t_{m_{11}}$	$t_{m_{12}}$
2*t_{q_1}		$h^1_{t_{q_1}}$	$h^2_{t_{q_1}}$	$h^3_{t_{q_1}}$								
				$y^{v_1}_{t_{q_1}}$	$y^{v_2}_{t_{q_1}}$	$y^{v_3}_{t_{q_1}}$						
2*t_{q_2}					$h^1_{t_{q_2}}$	$h^2_{t_{q_2}}$	$h^3_{t_{q_2}}$					
							$y^{v_1}_{t_{q_2}}$	$y^{v_2}_{t_{q_2}}$	$y^{v_3}_{t_{q_2}}$			
2*t_{q_3}								$h^1_{t_{q_3}}$	$h^2_{t_{q_3}}$	$h^3_{t_{q_3}}$		
										$y^{v_1}_{t_{q_3}}$	$y^{v_2}_{t_{q_3}}$	$y^{v_3}_{t_{q_3}}$
2*t_{q_4}	$h^3_{t_{q_4}}$										$h^1_{t_{q_4}}$	$h^2_{t_{q_4}}$
	$y^{v_1}_{t_{q_4}}$	$y^{v_2}_{t_{q_4}}$	$y^{v_3}_{t_{q_4}}$									

content in the monthly variables, they are transformed into three distinct quarterly series, as outlined in Castle *et al.* (2009). For every $\mathbf{z}_{t_m} = (z_{t_m}, \ldots, z_0)'$, denote the latest quarter t_q for which an observation $\mathbf{z}_{t_q}$ is available by τ, such that $\tau = t_q \leq t_m$, then:

$$\begin{aligned} \text{first month: } \mathbf{z}^1_{t_q} &= z_{\tau-2}, z_{\tau-5}, \ldots \\ \text{second month: } \mathbf{z}^2_{t_q} &= z_{\tau-1}, z_{\tau-4}, \ldots \\ \text{third month: } \mathbf{z}^3_{t_q} &= z_{\tau}, z_{\tau-3}, \ldots \end{aligned} \tag{78}$$

where $\mathbf{z}^1_{t_q}$, $\mathbf{z}^2_{t_q}$ and $\mathbf{z}^3_{t_q}$ are vectors containing observations for the first, second and third months of the quarter respectively. This transformation results in a total of 180 quarterly leading indicators, in contrast with the approach of aggregating monthly data into quarterly indicators: see for instance Ferrara *et al.* (2010).

To see how the availability of the quarterly indicators changes with respect to the nowcasting horizon, table 5 reports the relative timing for a particular example of the Q3 nowcasts at three different horizons, where leading indicator z_{t_q} is available with a lag of 3 months.

Table 5: Earliest available quarterly indicator relatively to nowcast horizons

Month	t_{m_5}	t_{m_6}	t_{m_7}	t_{m_8}	t_{m_9}	$t_{m_{10}}$	$t_{m_{11}}$	$t_{m_{12}}$
Horizon				$h^1_{t_{q_3}}$	$h^2_{t_{q_3}}$	$h^3_{t_{q_3}}$		
$h^1_{t_{q_3}}$	$z^2_{t_{q_3}-1}$							
$h^2_{t_{q_3}}$	$z^2_{t_{q_3}-1}$	$z^3_{t_{q_3}-1}$						
$h^3_{t_{q_3}}$	$z^2_{t_{q_3}-1}$	$z^3_{t_{q_3}-1}$	$z^1_{t_{q_3}}$					

The farthest available observation to the right in each row shows the last observation available at each nowcast horizon and corresponds to the release month shown in the top row. For instance, at $h^1_{t_{q_3}}$, when the first nowcast for Q3 is produced at t_{m_8}, the indicator observations are available until time $t_{m_8} - 3 = t_{m_5}$. This corresponds to the quarterly series $z^2_{t_{q_3}-1}$, which is observed in the second month of Q2. For the last horizon, respectively, the latest available observation is $z^1_{t_{q_3}}$. A similar lower triangular release structure is applicable to variables reported with a 1 and 2 month lags. Finally, the bridge equations for the quarterly GDP nowcasts are formed using the transformed quarterly series, which effectively solves the mixed frequency problem.

7.4 Model specifications

We now undertake a quasi real-time nowcasting exercise. All estimated models contain information that would have been available at the end of the month corresponding to each nowcast origin. For every horizon and every quarter, models are selected recursively using *Autometrics* with IIS at a 1% significance level, producing eight quarterly specifications for each subsample. Note that the nowcasting exercise is recursive in relation to the completeness of the information set. At each stage, the previous model is extended by additional data. Therefore, each new GUM nests the previous one. Each model is constructed using the transformed quarterly series for the leading indicators, their lags and a full set of impulse variables. This implies that the number of candidate variables exceeds the number of available observations. *Autometrics* can efficiently handle that issue by using an efficient tree search algorithm for model selection. The automatic model selection is then performed for the following specifications.

Vintage models (VM): a benchmark univariate model that includes the quarterly data on real time GDP growth vintages. In addition to the most accurate data at a real vintage v_3 this specification includes the earlier estimates, thereby utilizing the information contained in the revision process. It is noted that although the vintages are highly collinear, *Autometrics* is able to handle correlated data with satisfactory selection and forecasting performance, as was demonstrated by the simulation exercise above. To utilize all information available in real time, GUM differs for three nowcasting horizons and are shown in equation (79):

$$\begin{aligned}
h^1_{t_q} : \Delta\hat{y}_{t_q} &= f\left(\Delta y^{v_1}_{t_q-1},\ldots,\Delta y^{v_1}_{t_q-4};\Delta y^{v_2}_{t_q-1},\ldots,\Delta y^{v_2}_{t_q-4};\Delta y^{v_3}_{t_q-2},\ldots,\Delta y^{v_3}_{t_q-4}\right)\\
h^2_{t_q} : \Delta\hat{y}_{t_q} &= f\left(\Delta y^{v_1}_{t_q-1},\ldots,\Delta y^{v_1}_{t_q-4};\Delta y^{v_2}_{t_q-1},\ldots,\Delta y^{v_2}_{t_q-4};\Delta y^{v_3}_{t_q-1},\ldots,\Delta y^{v_3}_{t_q-4}\right)\\
h^3_{t_q} : \Delta\hat{y}_{t_q} &= f\left(\Delta y^{v_1}_{t_q},\ldots,\Delta y^{v_1}_{t_q-4};\Delta y^{v_2}_{t_q-1},\ldots,\Delta y^{v_2}_{t_q-4};\Delta y^{v_3}_{t_q-1},\ldots,\Delta y^{v_3}_{t_q-4}\right)
\end{aligned} \tag{79}$$

Note that at h^1_t , the most recent estimate of $\Delta y^{v_3}_{t_q-1}$ is still not available in real time and is therefore excluded from the GUM. Finally the last horizon h^3_t overlaps with the release period for the flash estimate and thus the first contemporaneous vintage of GDP is included in the GUM.

Single-indicator in-sample models (SM): for preliminary assessment of the predictors individual relevance, growth is projected onto the VM models each augmented with one single monthly indicator z^j_{i,t_q}, resulting in a total of 60 models for each evaluation period, corresponding to every leading indicator in the information set. Since no block separation is necessary here, the forecasts for the missing observation is omitted, utilizing in-sample information only. The maximum lag for the indicators is 12 and the minimum lag corresponds the earliest observed realization. Therefore the GUM for single models varies depending on the nowcasting horizon and the corresponding availability of the observations. The relevant lags are then selected by *Autometrics*. Since these models should merely provide an indication of individual predictors' robustness and relevance for nowcasting growth, the actual results for these models will not be presented, and instead assessed relatively to the VM model.

Augmented model with in-sample indicators (AI): the growth vintages from (79) are augmented with the full set of leading indicators. Only in-sample information enters the GUM so that the ragged edge problem is not corrected for. The GUM for the AI model is effectively a combination of the VM and SM specifications. The nowcasting model from the selected specification then has the following form:

$$\Delta\hat{y}^{h^k_q}_{t_q} = \hat{\boldsymbol{\alpha}}'\Delta\hat{\mathbf{y}}^v_{t_q} + \hat{\boldsymbol{\beta}}'\hat{\mathbf{z}}_{t_q} + \hat{\boldsymbol{\zeta}}'\hat{\mathbf{d}}_{t_q} \tag{80}$$

where $\Delta\hat{\mathbf{y}}_{t_q}^{v}$ is a vector containing the selected lags of the growth estimates, $\hat{\mathbf{z}}_{t_q}$ is a vector of lags of the relevant in-sample leading indicators and $\hat{\mathbf{d}}_{t_q}$ are significant impulse dummies.

Augmented models with forecasted indicators (AF): this model utilizes block separation and corrects the ragged edge problem by using forecasts for the missing monthly observations from (77) obtained from the blocks of correlated variables. The nowcasting models are then selected by *Autometrics* with IIS recursively using in-sample information on the indicators only. Based on that specification, a one-quarter ahead nowcast is produced by using in-sample observations and forecasted values if the selected model contains lags of leading indicators not yet available. A nowcasting model then has the following form:

$$\Delta\widehat{y}_{t_q}^{h_q^k} = \widehat{\boldsymbol{\alpha}}'\Delta\widehat{\mathbf{y}}_{t_q}^{v} + \widehat{\boldsymbol{\beta}}'\widehat{\mathbf{z}}_{t_q} + \widehat{\boldsymbol{\gamma}}'\widetilde{\mathbf{z}}_{t_q} + \widehat{\boldsymbol{\zeta}}'\widehat{\mathbf{d}}_{t_q} \tag{81}$$

where all variables are the same as in (80) and $\widetilde{\mathbf{z}}_{t_q}$ are the forecasted values for the selected relevant leading indicators. Comparison between AI and AF should reveal potential benefits from forecasting monthly indicators as compared to nowcasting from in-sample information only.

Augmented models with robust forecasts for indicators (AR): instead of the conditional monthly forecasts from (77), a differencing device is used for individual indicators forecasts that is robust to breaks, such that:

$$\widetilde{z}_{i,t_q} = z_{i,t_q-l_i} \tag{82}$$

where l_i is the release lag for variable i. The same type of GUM is used as for AI and AF and the nowcasting equation will include both in-sample and forecasting indicators similarly to (81). Note that the selected models for AR and AF will be identical, with the difference in the nowcasted values driven by the methodologies of indicator forecasts.

7.5 Vintage models

The nowcasts from VM for two subsamples are presented in figure 20, which plots the actual GDP growth against the nowcasts at three horizons respectively. Notice that for the first subsample, the nowcasts are identical for all three horizons, since the selected models do not differ across the horizons. In fact, *Autometrics* did not select any vintages as relevant and the final specification only includes an intercept, which is the same for all three horizons. Consequently, preliminary vintages do not have any explanatory power for the latest estimate of the GDP growth between 1992 and 2000. The conditional in-sample mean produces the best parsimonious nowcasting model. This is in line with the results obtained in Giannone *et al.* (2008), who report that during stable growth periods the conditional means are difficult to beat on average.

Subsample 2 is characterize by negative growth for 5 consecutive quarters starting from 2008Q2. For this and the subsequent quarter, VM predicts the output to grow. The vintages model nowcasts the start of the recession two quarters after its start, in 2008Q4, and from then on all three vintages can track the series quite closely. This finding, once again, underlines the weakness of the preliminary estimates and their lags in explaining actual growth. It is stressed that at all horizons the nowcasted growth paths are quite similar. Moreover, the model specification at $h_{t_q}^3$, when the first preliminary vintage is released, provides a noticeable improvement for most quarters, as compared to the earlier horizons. Note, however, that the first recovery period in 2009Q3 is correctly predicted at all horizons, but the magnitude of the error is quite significant. Still, a promising finding is that the model from horizon 3 that includes $\Delta y_{t_q}^{v_1}$ in the selected specification have the lowest RMSE, which provides some reassurance for applying contemporaneous information to improve the nowcasting accuracy.

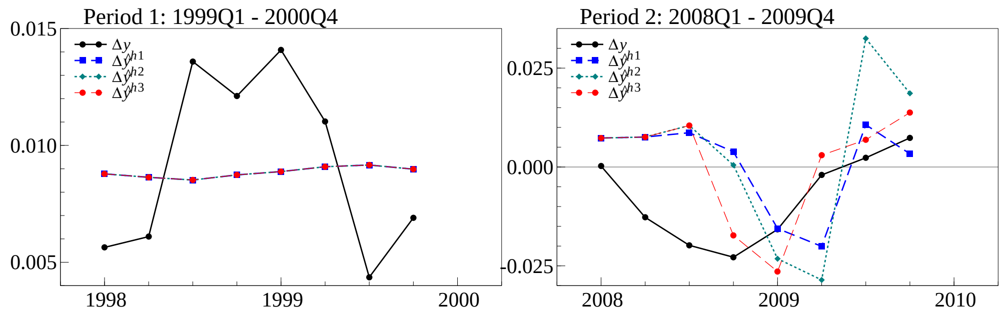


Figure 20: Nowcasts from vintages model for three horizons and two subsamples: 1999Q1–2000Q4 (left panel) and 2008Q1–2009Q4 (right panel)

7.6 Single-indicator models

The single indicator models are compared to the VM model in terms of the relative RMSE and MAE for the two periods. Figure 21 depicts the ratio for the single models RMSE to the vintages model RMSE and a similar ratio for MAE. Each point corresponds to one single indicator model. On the x-axis the ratio for the first subsample is plotted and y-axis shows the ratio for the second subsample. If the ratio is smaller than 1, the model performs better than VM for that particular period. A point that lies within the green square corresponds to a model for an indicator that provides consistent improvement relatively to VM in both subsamples. The area marked by a red square corresponds to robustly underperforming indicators.

From the pattern emerging in figure 21, there is little evidence of robustly performing models across the set of indicators. The latter effect would be apparent if the points clustered around the 45 degree line starting from the origin. In fact, there is only a handful of models that provide improvement on VM for every horizon. For instance, at horizon 2 only three indicators perform better in terms of RMSE and MAE. For horizon 2, however, MAE draws a more promising picture, with around half of the indicators being able to improve on the VM models for the second subsample. Out of these models, 15 of which are also better predictors of GDP growth in period 1. For horizon 2, the models are least disperse. However, for horizon 3, only a few single models are again robustly better than VM, while improved performance for the second period is noticed for the majority of the models.

Only three variables have shown a robust accuracy improvement across all periods, nowcast horizons and accuracy measures. These are the consumer confidence index, non-farm payrolls and ICS housing survey, all corresponding to block 1. The nowcasting output for the single models corresponding to these three variables together with the VM models for the second subsample are presented in figure 22. The main feature is that the SM models can nowcast the recession one quarter before VM at almost all horizons, while keeping nowcasting errors lower.

Overall, there is little indication of consistent model performance for the single models. The number of indicators that overperform VM models varies dramatically across subsamples, nowcasting horizons and nowcasting accuracy measures. In addition, there is no robust evidence that some indicators are completely irrelevant either. Nevertheless, some single indicators can deliver nowcasting improvement for all horizons and subsamples, implying that the chosen information set is not irrelevant and that *Autometrics* will be able to produce accurate nowcasts when the indicators are combined.

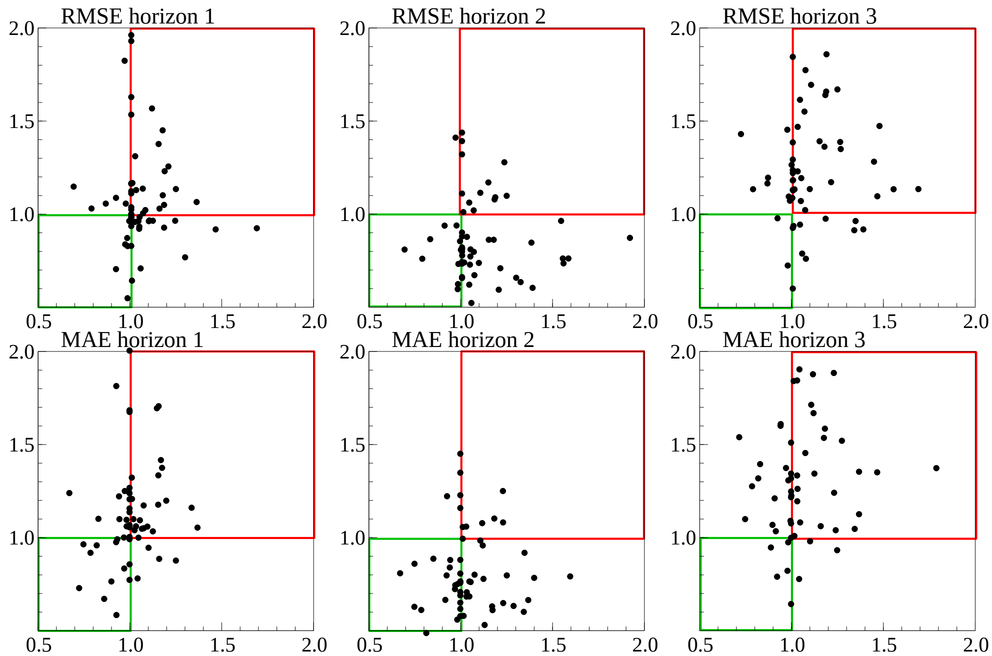


Figure 21: Scatter plot of RMSE (top three plots) and relative MAE (bottom three plots) for vintages models augmented by single indicators relative to vintages model for two periods and three nowcasting horizons. Ratio of RMSE for single model to vintages model for the first period, 1992–2000, on the x-axis and for the second period, 2001–2009, on the y-axis for the three nowcasting horizons. Red area defines the region of robust underperformance, and green, a region of robust overperformance

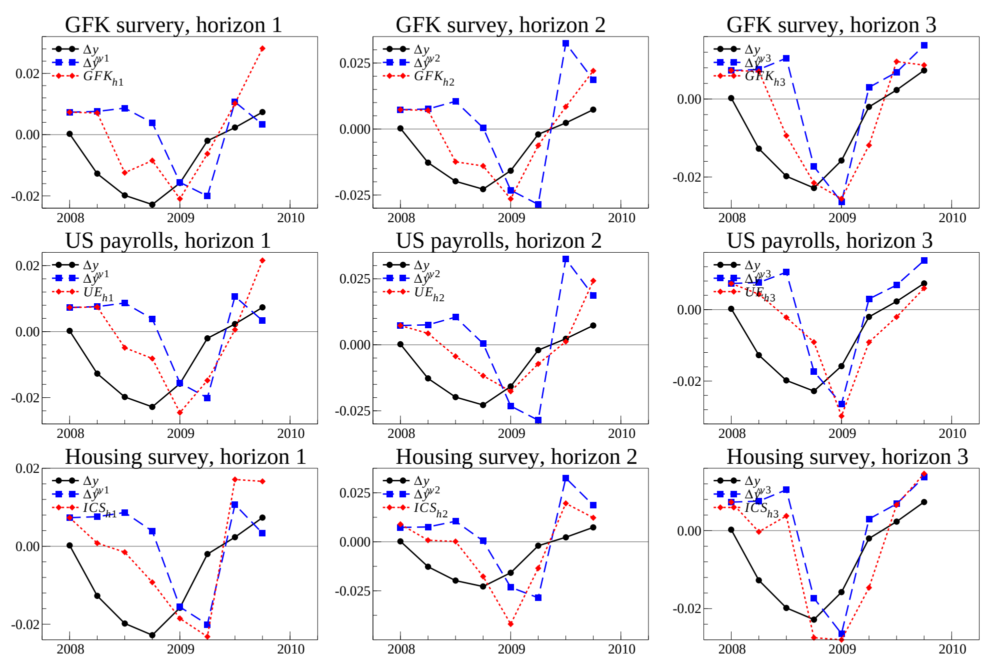


Figure 22: Comparing VM and SM outcomes

7.7 Augmented models

Three versions of the models augmented with all 60 monthly leading indicators are compared to the vintages model. Table 6 presents RMSE and MAE for the two subsamples and three nowcasting horizons, where the reported values for the augmented models represent the ratios to RMSE and MAE of the vintages model, respectively. First, note that the VM model performs much better on average in period 1. The reported RMSE and MAE are higher between 2001 and 2009. This is expected from the previously shown results, namely that VM for the first period is essentially estimating the conditional mean, since only a constant is retained in the final models. However, high volatility and the changing signs of the series in the second subsample are not very well captured by the growth estimates alone.

	RMSE					
	1993 - 2000			2001 - 2009		
Model	$h^1_{t_q}$	$h^2_{t_q}$	$h^3_{t_q}$	$h^1_{t_q}$	$h^2_{t_q}$	$h^3_{t_q}$
Vintages	*0.004*	*0.004*	*0.004*	*0.017*	*0.022*	*0.014*
Augmented in-sample*	1.178	1.868	1.306	0.691	0.556	0.664
Augmented forecasts*	1.442	1.029	1.142	0.791	0.438	0.689
Augmented robust*	1.476	1.208	1.195	0.784	0.438	0.692

*RMSE ratio to RMSE of the corresponding vintages model

	MAE					
	1993 - 2000			2001 - 2009		
Model	$h^1_{t_q}$	$h^2_{t_q}$	$h^3_{t_q}$	$h^1_{t_q}$	$h^2_{t_q}$	$h^3_{t_q}$
Vintages	*0.004*	*0.004*	*0.004*	*0.014*	*0.020*	*0.011*
Augmented in-sample*	1.014	1.605	1.111	0.686	0.489	0.754
Augmented forecasts*	1.230	0.906	0.996	0.818	0.431	0.788
Augmented robust*	1.294	1.028	1.032	0.793	0.426	0.769

*MAE ratio to MAE of the corresponding vintages model

Table 6: RMSE of augmented models as ratio to RMSE of the vintages model (top table), MAE of augmented models as ratio to MAE of the vintages model (bottom table)

The same outcome for the relative accuracy between the subsamples is found for all augmented models. In addition, a particularly interesting finding for the first subsample is that augmenting the model by leading indicators tends to reduce accuracy on average. It is not the case only for the AG model in horizons 2 and 3, for which the monthly variables slightly improve MAE of the nowcasts. In fact, the models augmented by indicators forecasts perform better than the other two augmented specifications for the later horizons. However, there is no consistent improvement as time elapses, since MSE does not reduce monotonically as the nowcast horizon moves closer to the first growth estimate release date.

For some models, nowcasts for horizon 2 are more accurate than those in horizon 3, whereas for others that relationship is reversed. In particular, models with the corrected ragged edges show consistently higher RMSE and MAE for the first horizon. This suggests that if the bridge equations framework is used to produce balanced panels by predicting the monthly indicators, inflow of disaggregated information tends to improve nowcasting accuracy. The opposite is true for the AS model as the most accurate nowcasts are produced in $h^1_{t_q}$. When juxtaposing AS with AR, the latter produce better growth estimates on average when some information about the last month of the quarter become available, i.e., horizons

2 and 3. The results for the stable positive growth subsample in general support previous findings for alternative nowcasting methodologies.

In terms of the particular variables selected by *Autometrics* in the final specification of the augmented models for the first subsample, growth rates are never retained. There is some robustness in the selected models for different nowcasting horizons, where the retained variables do not vary considerably among specifications. On average, around 10 lags are selected in each specification. There is much more variability in terms of the selected variables and their lags across nowcasting origins, meaning that there is little robustness across time periods. The variables that were most often retained are the US consumer sentiment index, UK industrial production of various products and the average discount rate of the UK treasury bills. In general, due to lack of consistency in the terminal augmented models and their inferior performance relatively to the vintages model, it is suggested that the leading indicators are not very good predictors of growth when it is stable. The latter effect becomes even more prominent when examining the actual nowcasts produced by the augmented models for the first subsample, depicted in figure 23 for three horizons. Significant differences can be noticed across the models and horizons.

The second subsample represents a more interesting case due to high variability and alternating directions of growth. Table 6 shows that unlike in the first subsample, all augmented models provide substantial improvement on the VM, as mirrored by RMSE and MAE ratios much lower than unity. In fact, horizon 2 nowcasts have the greatest relative accuracy enhancement, with both RMSE and MAE reduced by a factor larger than 2. At horizons 1 and 3, the ratios are slightly higher. Note, however, that horizon 3 has a much lower levels RMSE and MAE for the vintages model induced by inclusion of the first flash contemporaneous growth estimate into the GUM, which is in fact selected by *Autometrics* as relevant at various nowcasting origins. Consequently, it is concluded that in absolute terms all augmented models produce progressively more accurate nowcasts as new disaggregated information becomes available.

Comparison of the augmented models relative performance reveals that in-sample models perform better for horizon 1 and 3. Although rather insignificant, this finding suggests that using *Autometrics* to forecast individual indicators and then using those forecasts in the nowcasting bridge equations actually reduces overall accuracy. Although true on average, when looking at point nowcasts presented in figure 23, it is apparent that AF and AR can track major turns in the evolution of the GDP growth series slightly better than the AS model. This effect is particularly noticeable in horizons 1 and 2.

One of the key results in this empirical exercise is the satisfactory performance of the robust device as applied to forecasting the leading indicators that face contemporaneous breaks. The model augmented with robust forecasts is, in fact, the one that produces the most timely nowcast of the first instance of output decline. AR specification correctly predicts negative growth for 2008Q2 at the second month of the quarter, i.e., two months before the release of the first ONS estimate. This finding confirms that a device robust to breaks can yield some improvement of the nowcasting accuracy. This provides further support for the results obtained in the simulations chapter.

Unlike the AR model, AS and AF nowcast the change of the sign in the growth series at the end of the quarter. During the recessionary period, all models provide relatively good estimations of GDP growth, which are much more accurate than the benchmark vintages model and the estimates published by the ONS.

When analyzing the specific variables and their lags retained in final specifications, only one leading indicator is consistently selected as relevant. This is the UK jobless claimants count rate, which is available with a one month delay and is chosen by all models at all horizons and nowcast origins. Various industrial production indexes for the UK and the US as well as the GFK consumer confidence index are also chosen in most models. This combination of hard and soft data was expected from their

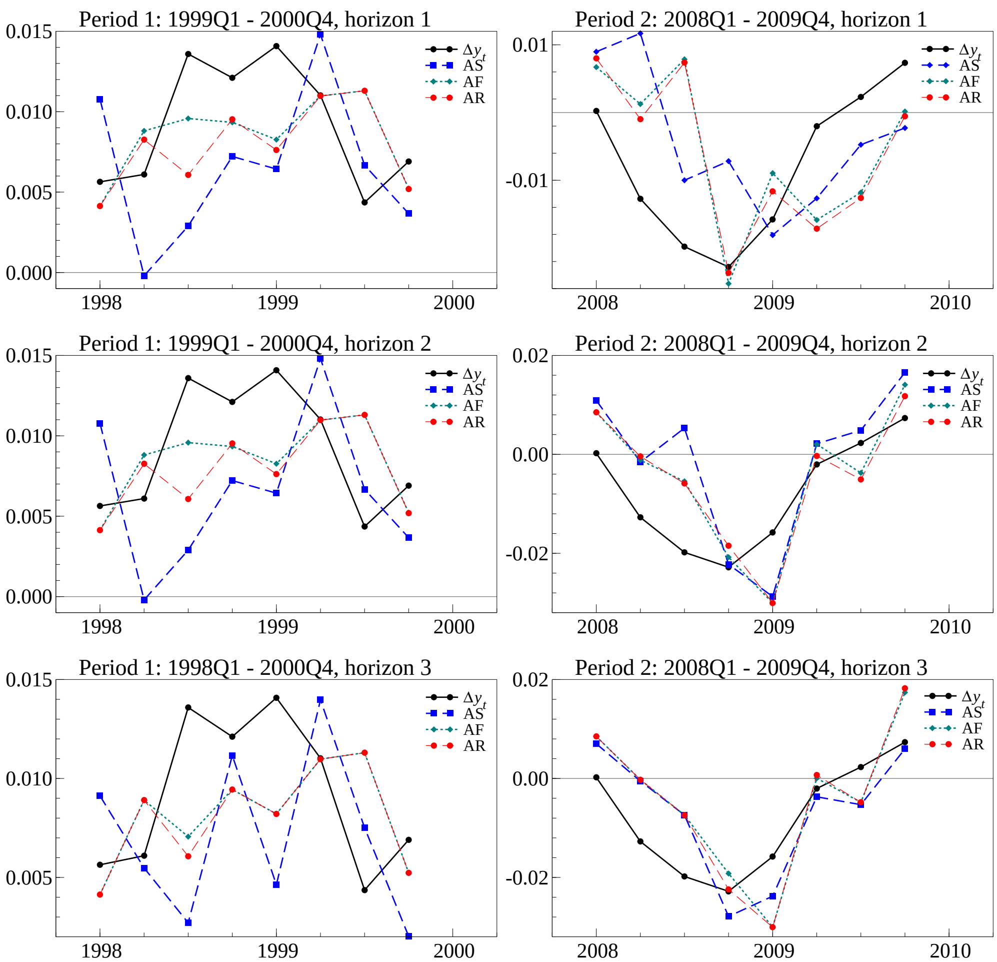


Figure 23: Nowcasts from augmented models for two subsamples: 1999Q1–2000Q4 (left panels) and 2008Q1–2009Q4 (right panels) and three horizons: horizon 1 (top panels), horizon 2 (middle panels), and horizon 3 (bottom panels)

robust performance in the single model specifications and confirms most previous findings. Finally, at all horizons, the financial variables in block 5, such as the FTSE index are occasionally selected. However, the lags vary substantially and the coefficients are almost negligible so have insignificant explanatory power for the growth series. All in all, there is some evidence of consistency in terms of the relevant variables, and very limited indications that the timing of the influence is robust, since the selected lags differ considerably across nowcasting origins and horizons.

7.8 Contemporaneous breaks

An examination of the contemporaneous impulse dummies retained in the forecasting models for the monthly indicators is conducted at a nominal 1% significance level and a much tighter 0.1%, suggested as the preferred strategy in Castle *et al.* (2009). Since the first subsample is characterized by low variance and mean reverting positive growth, there were no breaks detected for the GDP series at either significance level. Moreover, the monthly indicators were in general quite stable with only a few impulse dummies detained at the 1% level and virtually none at 0.1%. Thus, any further break analysis for the first subsample seems irrelevant.

In the second subsample, on the other hand, multiple are breaks detected for the growth variable. In particular, in-sample breaks are significant at 0.1% level for 2008Q2, 2008Q3 and 2008Q4, the periods of the largest output declines. This confirms that the recession was a location shift relative to previous information for the GDP growth series. Thus, the ability of the augmented nowcasting models to produce good predictions of the output decline further confirms that disaggregated variables contain timely and relevant information to predict major macroeconomic movements, even when the latter have unanticipated and highly significant location shifts.

An analysis of the common contemporaneous breaks was conducted within 6 blocks of correlated variables for the second subsample, and the second quarter of 2008 in particular. At the 1% level, too many breaks were detected to be useful, whereas when a tighter significance level at 0.1% is used, few dummies are retained. However, during the months of 2008Q2, those variables that were retained in the terminal nowcasting models at each horizon do not contain any contemporaneous breakst. On the contrary, variables that do break during these months were not selected. These results provide insufficient evidence for the contemporaneous intercept device to be applied for the correction of the monthly indicators' forecasts. Alternative methods of blocking variables could be examined: testing for cointegration or common serial correlation could reveal links between disaggregated variables.

8 Some conclusions

First, some conclusions on model-based forecasting in non-stationary economies subject to unanticipated structural breaks, where models differ from DGPs in unknown ways, selected from possibly unreliable data. We showed that the forecasting implications differ considerably from a setting where the model coincided with the DGP in a constant mechanism. That was because unanticipated location shifts are pernicious for forecasting, leading to systematic mis-forecasting in all forms of equilibrium-correction models. Conversely, if no location shift occurred, every DGP parameter could be shifted without any noticeable failure.

That led us to conclude that robust devices could play an invaluable role after location shifts occurred, albeit that they would not forecast such breaks. We explained why the pre-existing conditional expectations, even given all relevant information, would not provide a good forecast when unanticipated location shifts occurred, which in turn invalidates inter-temporal theory derivations based on the law of iterated expectations. As almost identical location shifts could be created by many different combinations of DGP parameters shifting, and which had changed may not be discernable till well after the event, economic agents as well as nowcasters and forecasters, would be unable to quickly learn how to foresee in the changed environment.

Nevertheless, systematic mis-forecasting could be mitigated by differencing the selected econometric system, retaining original estimates, even if the DGP parameters changed. Devices like $\Delta^2\widehat{\mathbf{x}}_{T+h|T+h-1} = \mathbf{0}$ are knowingly mis-specified in-sample by restricted information, yet avoid systematic forecast failure on forecasts made after a shift. Moreover, the costs of unnecessary differencing when there is no shift are relatively small compared to the costs of ignoring a break when it has happened. Such results have implications for modeling methodology, as they stress that forecast failure requires a location shift in the DGP relative to the empirical model, and other changes in the DGP may have little observable impact. Thus, the verisimilitude of a model cannot be checked by forecasting success or failure.

Our proposed approach uses *Autometrics* to select over many potentially relevant covariates using higher frequency data, with in-sample and end-point location shift detection based on IIS. Such equations then forecast all disaggregates ready for comparison with incoming measured disaggregates. Forecasts of the missing disaggregates can then be adjusted when there is evidence of location shifts occurring within the period. Alternatively, the robust devices could be used. No device considered here can forecast future location shifts. A different class of model is needed for that, based on different information (see e.g., Castle *et al.*, 2011).

Nowcasting has much in common with, but also differs substantively from, forecasting due to the different roles played by problems of missing data, measurement errors, recording delays, changing databases, and location shifts. The apparent 'break down' of the UK ONS's current methods, noted in the report by the *Financial Times* above, shows the difficulty of nowcasting in times of economic uncertainty and structural change. Location shifts induce nowcast failure, but interact with measurement errors to make discrimination between the causes difficult within the available time horizon for aggregates, although discrimination seems a lesser problem when building an aggregate from large numbers of disaggregates.

The principle findings of the empirical nowcasting exercise for UK quarterly GDP growth confirm that models augmented with disaggregated monthly variables and selected by *Autometrics* with impulse-indicator saturation have superior predictive power to the official estimates and the univariate benchmark models. The relative nowcasting accuracy is further enhanced during the period spanning five consecutive quarters of negative growth, for which highly significant breaks in the series were detected. In addition, it is found that inflow of disaggregated data throughout the reference quarter tends to improve absolute nowcasting accuracy monotonically, irrespective of the information content in the augmented models. On the other hand, the augmented model with contemporaneous break detection, and subsequent robust correction for the leading indicators' forecasts, produces the most accurate nowcasts. In particular, this model predicts the first quarter of output decline in 2008 one month in advance. For comparison, models that use in-sample or forecast conditioning variables without explicitly modeling breaks were able to predict the same event only after the end of the reference quarter. These findings provide substantial evidence that *Autometrics* model selection with IIS can be effectively used for nowcasting GDP from disaggregated data even when contaminated by multiple structural breaks.

References

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M., and Veronese, G. (2010). New EuroCOIN: Tracking economic growth in real time. *Review of Economics and Statistics*, **92**, 1024–1034.

Andreou, E., Ghysels, E., and Kourtellos, A. (2011). Forecasting with mixed-frequency data. In Clements, and Hendry (2011), pp. 225–246.

Angelini, E., Camba-Méndez, G., Giannone, D., Rünstler, G., and Reichlin, L. (2011). Short-term forecasts of Euro-area GDP growth. *Econometrics Journal*, **14**, C25–C44.

Angelini, E., Henry, J., and Marcellino, M. (2006). Interpolation and backdating with a large information set. *Journal of Economic Dynamics and Control*, **30**, 2693–2724.

Artis, M., Banerjee, A., and Marcellino, M. (2005). Factor forecasts for the UK. *Journal of Forecasting*, **24**, 279–298.

Bánbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting. In Clements, and Hendry (2011), Ch. 7.

Bańbura, M., and Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, **27**, 333–346.

Banerjee, A., Marcellino, M., and Masten, I. (2058). Leading indicators for Euro area Inflation and GDP Growth. *Oxford Bulletin of Economics and Statistics*, **67**, 785–813.

Bardsen, G., Hurn, S., and McHugh, Z. (2010). Asymmetric unemployment rate dynamics in Australia. Creates research paper 2010-2, Aarhus University, Denmark.

Barhoumi, K., Benk, S., Cristadoro, R., Reijer, A. D., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., Van Nieuwenhuyze, C., and Rünstler, G. (2008). Short-term forecasting of GDP using large monthly datasets–A pseudo real-time forecast evaluation exercise. Working paper 133, National Bank of Belgium, Brussels.

Barrell, R. (2001). Forecasting the world economy. In Hendry, D. F., and Ericsson, N. R. (eds.), *Understanding Economic Forecasts*, pp. 149–169. Cambridge, Mass.: MIT Press.

Barrell, R., Dury, K., Holland, D., Pain, N., and te Velde, D. (1998). Financial market contagion and the effects of the crises in East Asia, Russia and Latin America. *National Institute Economic Review*, **166**, 57–73.

Boivin, J., and Ng, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, **1(3)**.

Bontemps, C., and Mizon, G. E. (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics*, **70**, 721–750.

Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day. First published, 1970.

Brisson, M., Campbell, B., and Galbraith, J. W. (2003). Forecasting some low predictability time series using diffusion indices. *Journal of Forecasting*, **22**, 515–531.

Camacho, M., and Sancho, I. (2003). Spanish diffusion indexes. *Spanish Economic Review*, **5**, 173–203.

Campos, J., Ericsson, N. R., and Hendry, D. F. (2005). Editors' introduction. In Campos, J., Ericsson, N. R., and Hendry, D. F. (eds.), *Readings on General-to-Specific Modeling*, pp. 1–81. Cheltenham: Edward Elgar.

Castle, J. L., Clements, M. P., and Hendry, D. F. (2012). Forecasting by factors, by variables, by both or neither?. Discussion paper 600, Economics Department, University of Oxford.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011a). Evaluating automatic model selection. *Journal of Time Series Econometrics*, **3 (1)**, DOI: 10.2202/1941–1928.1097.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012a). Model selection in equations with many 'small' effects. *Oxford Bulletin of Economics and Statistics*, DOI: 10.1111/j.1468-0084.2012.00727.x

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012b). Model selection when there are multiple breaks. *Journal of Econometrics*, **169**, 239–246.

Castle, J. L., Doornik, J. A., Hendry, D. F., and Nymoen, R. (2013). Testing the invariance of expectations models of inflation. Forthcoming *Econometric Reviews*.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2009). Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review*, **210**, 71–89.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2010). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, **158**, 25–36.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2011). Forecasting Breaks and During Breaks. In Clements, and Hendry (2011), pp. 315–353.

Castle, J. L., and Hendry, D. F. (2010). Nowcasting from disaggregates in the face of location shifts. *Journal of Forecasting*, **29**, 200–214.

Castle, J. L., and Hendry, D. F. (2011). Automatic selection of non-linear models. In Wang, L., Garnier, H., and Jackman, T. (eds.), *System Identification, Environmental Modelling and Control*, pp. 229–250. New York: Springer.

Castle, J. L., and Hendry, D. F. (2012). Model selection in under-specified equations with breaks. *Journal of Econometrics*, forthcoming.

Castle, J. L., and Shephard, N. (eds.)(2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Choi, H., and Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, **88**, 2–9.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591–605.

Clements, M. P., and Galvão, A. B. (2008). Macroeconomic forecasting with mixed frequency data: Forecasting US output growth. *Journal of Business and Economic Statistics*, **26**, 546–554.

Clements, M. P., and Galvão, A. B. (2009a). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, **24**, 1187–1206.

Clements, M. P., and Galvão, A. B. (2009b). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, **24**, 1187–1206.

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M. P., and Hendry, D. F. (2001a). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, **17**, 550–554.

Clements, M. P., and Hendry, D. F. (2001b). An historical perspective on forecast errors. *National Institute Economic Review*, **177**, 100–112.

Clements, M. P., and Hendry, D. F. (eds.)(2002). *A Companion to Economic Forecasting*. Oxford: Blackwells.

Clements, M. P., and Hendry, D. F. (2003). Forecasting in the National Accounts at the Office for National Statistics. Report no 12, Statistics Commission.

Clements, M. P., and Hendry, D. F. (2005). Guest Editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics*, **67**, 713–753.

Clements, M. P., and Hendry, D. F. (2008). Economic forecasting in a changing world. *Capitalism and Society*, **3**, 1–18.

Clements, M. P., and Hendry, D. F. (eds.)(2011). *Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press.

Corradi, V., Fernandez, A., and Swanson, N. R. (2009). Information in the revision process of real-time datasets. *Journal of Business and Economic Statistics*, **27**, 455–467.

Croxson, K., and Reade, J. J. (2008). Information and efficiency: Goal arrival in soccer betting. Mimeo, Economics Department, Oxford University.

D'Agostino, A., and Giannone, D. (2012). Comparing alternative predictors based on large-panel factor models. *Oxford Bulletin of Economics and Statistics*, **74**, 306–326.

D'Agostino, A., McQuinn, K., and O'Brien, D. (2011). Nowcasting irish GDP. MPRA paper 32941, University of Munich, Germany.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?. *Journal of Econometrics*, **146**, 318–328.

den Reijer, A. (2005). Forecasting Dutch GDP using Large Scale Factor Models. Working paper 028, Netherlands Central Bank, Amsterdam.

Diebold, F. X., and Rudebusch, G. D. (1991). Turning point prediction with the composite leading index: An ex ante analysis. In Lahiri, K., and Moore, G. H. (eds.), *Leading economic indicators: New approaches and forecasting records*, pp. 231–256. Cambridge: Cambridge University Press.

Diron, M. (2008). Short-term forecasts of Euro-area real GDP growth: An assessment of real-time performance based on vintage data. *Journal of Forecasting*, **27**, 371–390.

Doornik, J. A. (2007). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.

Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics*, **70**, 915–925.

Doornik, J. A. (2009a). Autometrics. In Castle, and Shephard (2009), pp. 88–121.

Doornik, J. A. (2009b). Improving the timeliness of data on influenza-like illnesses using Google Trends. Typescript, Department of Economics, University of Oxford.

Doornik, J. A., and Hendry, D. F. (2009). *Empirical Econometric Modelling using PcGive: Volume I*. London: Timberlake Consultants Press.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, **164**, 188–205.

Eickmeier, S., and Ziegler, C. (2008). How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *Journal of Forecasting*, **27**, 237–265.

Elliott, G., Granger, C. W. J., and Timmermann, A. (eds.)(2006). *Handbook of Econometrics on Forecasting*. Amsterdam: Elsevier.

Emerson, R. A., and Hendry, D. F. (1996). An evaluation of forecasting using leading indicators. *Journal* *of Forecasting*, **15**, 271–91.

Engle, R. F., and Hendry, D. F. (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics*, **56**, 119–139.

Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304.

Faust, J., Rogers, J. H., and Wright, J. H. (2007). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking*, **37**, 403–419.

Ferrara, L., Guegan, D., and Rakotomarolahy, P. (2010). GDP nowcasting with ragged-edge data: A semi-parametric modelling. *Journal of Forecasting*, **29**, 186–199.

Fildes, R., and Ord, K. (2002). Forecasting competitions – their role in improving forecasting practice and research. In Clements, and Hendry (2002), pp. 322–253.

Fildes, R. A., and Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*, **63**, 289–308.

Fixler, D. J., and Grimm, B. T. (2008). The reliability of the GDP and GDI estimates. *Survey of Current Business*, **88**, 16–32.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2001). Coincident and leading indicators for the Euro Area. *Economic Journal*, **111**, 62–85.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the Euro Area?. *Journal of Monetary Economics*, **50**, 1243–1255.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, **100**, 830–840.

Frisch, R. (1938). Statistical versus theoretical relations in economic macrodynamics. Mimeograph dated 17 July 1938, League of Nations Memorandum. Reproduced by University of Oslo in 1948 with Tinbergen's comments. Contained in Memorandum 'Autonomy of Economic Relations', 6 November 1948, Oslo, Universitets Økonomiske Institutt.

Frisch, R., and Waugh, F. V. (1933). Partial time regression as compared with individual trends. *Econometrica*, **1**, 221–223.

Gertler, M., and Rogoff, K. (eds.)(2005). *NBER Macroeconomics Annual 2004*. Cambridge Mass.: MIT Press.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The MIDAS touch: MIxed DAta Sampling regression models. mimeo, Chapel Hill, N.C.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, **26**, 53–90.

Giacomini, R., and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, **74**, 1545–1578.

Giannone, D., Reichlin, L., and Simonelli, S. (2009). Nowcasting Euro Area economic activity in Real-Time: The role of confidence indicators. *National Institute Economic Review*, **210**, 90–97.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting GDP and Inflation: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, **55**, 665–676.

Gil, R., and Levitt, S. D. (2007). Testing the efficiency of markets in the 2002 world cup. *Journal of Prediction Markets*, **1**, 255–270.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, **12**, 1–118.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics*, **11**, 45–65.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.

Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In Mills, T. C., and Patterson, K. D. (eds.), *Palgrave Handbook of Econometrics*, pp. 3–67. Basingstoke: Palgrave MacMillan.

Hendry, D. F., and Clements, M. P. (2004). Pooling of forecasts. *Econometrics Journal*, **7**, 1–31.

Hendry, D. F., and Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics*, **29**, 216–227.

Hendry, D. F., and Johansen, S. (2012). Model discovery and Trygve Haavelmo's legacy. *Econometric Theory*, forthcoming.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, **33**, 317–335. Erratum, 337–339.

Hendry, D. F., and Krolzig, H.-M. (2003). New developments in automatic general-to-specific modelling. In Stigum, B. P. (ed.), *Econometrics and the Philosophy of Economics*, pp. 379–419. Princeton: Princeton University Press.

Hendry, D. F., and Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, **115**, C32–C61.

Hendry, D. F., and Mizon, G. E. (2010). On the mathematical basis of inter-temporal optimization. Discussion paper 497, Economics Department, Oxford.

Hendry, D. F., and Mizon, G. E. (2011a). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, **3 (1)**, DOI: 10.2202/1941–1928.1100.

Hendry, D. F., and Mizon, G. E. (2011b). What needs re-thinking in macroeconomics?. *Global Policy*, **2**, 176–183.

Hendry, D. F., and Mizon, G. E. (2012a). Forecasting from structural econometric models. Working paper 597, Economics Department, Oxford University.

Hendry, D. F., and Mizon, G. E. (2012b). Open-model forecast-error taxonomies. In Chen, X., and Swanson, N. R. (eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pp. 219–240. New York: Springer.

Hendry, D. F., and Pretis, F. (2012). Anthropogenic influences on atmospheric CO2. In Fouquet, R. (ed.), *Handbook on Energy and Climate Change*, forthcoming. Cheltenham: Edward Elgar.

Hendry, D. F., and Reade, J. J. (2008a). Economic forecasting and prediction markets. Working paper, Economics Department, Oxford University.

Hendry, D. F., and Reade, J. J. (2008b). Modelling and forecasting using model averaging. Working paper, Economics Department, Oxford University.

Hendry, D. F., and Richard, J.-F. (1989). Recent developments in the theory of encompassing. In Cornet, B., and Tulkens, H. (eds.), *Contributions to Operations Research and Economics. The XXth Anniversary of CORE*, pp. 393–440. Cambridge, MA: MIT Press.

Hendry, D. F., and Santos, C. (2010). An automatic test of super exogeneity. In Watson, M. W., Bollerslev, T., and Russell, J. (eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford: Oxford University Press.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. ONR Research Memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

Hubrich, K. (2005). Forecasting Euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?. *International Journal of Forecasting*, **21**, 119–136.

Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, and Shephard (2009), pp. 1–36.

Kapetanios, G., and Marcellino, M. (2010). Factor-GMM estimation with large sets of possibly weak instruments. *Computational Statistics and Data Analysis*, **54**, 2655–2675.

Kitchen, J., and Monaco, R. M. (2003). Real-time forecasting in practice: the U.S. Treasury staff's real-time GDP forecast system. *Business Economics*, **October**, 10–19.

Kitov, O. I. (2012). Nowcasting GDP from disaggregated data in the presence of structural breaks. Unpublished M.Phil. Thesis, University of Oxford.

Kuzin, V., Marcellino, M., and Schumacher, C. (2009a). Pooling versus model selection for nowcasting with many predictors: An application to German GDP. Discussion paper series 1: Economic studies, no. 3, Deutsche Bundesbank.

Kuzin, V., Marcellino, M., and Schumacher, C. (2009b). Pooling versus model selection for nowcasting with many predictors: Empirical evidence for six industrial countries. Economics working paper,eco2009/13, European University Institute.

Lee, K., Olekalns, N., and Shields, K. (2008). Nowcasting, business cycle dating and the interpretation of new information when real time data are available. Discussion papers in economics 08/17, University of Leicester.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In Brunner, K., and Meltzer, A. (eds.), *The Phillips Curve and Labor Markets*, Vol. 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19–46. Amsterdam: North-Holland Publishing Company.

Lütkepohl, H. (2006). Forecasting with VARMA models. In Elliott *et al.* (2006), pp. 288–325.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., *et al.* (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, **1**, 111–153.

Makridakis, S., and Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–476.

Marcellino, M., and Schumacher, C. (2010). Factor-MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, **72**, 518–550.

Mariano, R., and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, **18**, 427–443.

Mayo, I., and Espasa, A. (2012). Forecasting aggregates and disaggregates with common features. Forthcoming, *International Journal of Forecasting*.

Mills, T. C. (2010). Bradford Smith: An econometrician decades ahead of his time. *Oxford Bulletin of Economics and Statistics*, **73**, 276–285.

Mitchell, J. (2009). Where are we now? The UK recession and nowcasting GDP growth using statistical models. *National Institute Economic Review*, **209**, 60–69.

Mitchell, J., Smith, R. J., Weale, M. R., Wright, S., and Salazar, E. (2005). An indicator of monthly GDP and an early estimate of quarterly GDP growth. *Economic Journal*, **115**, F108–F129.

Mizon, G. E., and Richard, J.-F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, **54**, 657–678.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1998). Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association*, **93**, 478–493.

Rapach, D. E., and Wohar, M. E. (eds.)(2008). *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Bingley, UK: Emerald Group.

Reed, G. (2000). How the preliminary estimate of GDP is produced. *Economic Trends*, **556**, 53–61.

Reed, G. (2002). How much information is in the UK preliminary estimate of GDP?. *Economic Trends*, **585**, 1–8.

Rünstler, G., and Sédillot, F. (2003). Short-term estimates of Euro area real GDP by means of monthly data. Working Paper series, no. 276, European Central Bank.

Salkever, D. S. (1976). The use of dummy variables to compute predictions, prediction errors and confidence intervals. *Journal of Econometrics*, **4**, 393–397.

Schumacher, C., and Breitung, J. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting*, **26**, 271–302.

Schumacher, C., and Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, **24**, 386–398.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Smith, B. B. (1927). Forecasting the volume and value of the cotton crop. *Journal of the American Statistical Association*, **22**, 442–459.

Smith, B. B. (1929). Judging the forecast for 1929. *Journal of the American Statistical Association*, **24**, 94–98.

Stock, J. H., and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, **14**, 11–30.

Stock, J. H., and Watson, M. W. (1998). Diffusion indices. Working paper, 6702, NBER, Washington.

Stock, J. H., and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97**, 1167–1179.

Stock, J. H., and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indices. *Journal of Business and Economic Statistics*, **20**, 147–162.

Stock, J. H., and Watson, M. W. (2006). Forecasting with many predictors. In Elliott *et al.* (2006), pp. 515–554.

Stone, J. R. N. (1947). On the interdependence of blocks of transactions. *Journal of the Royal Statistical Society*, **8**, 1–32. Supplement.

Wallis, K. F. (1986). Forecasting with an econometric model: The 'ragged edge' problem. *Journal of Forecasting*, **5**, 1–14.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, **6**, 324–342.

Wolfers, J., and Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, **18**, 107–126.